The Origin of Memory for Language

Silvia Benavides Varela

Thesis submitted for the degree of *Doctor Philosophiae* in
Cognitive Neurosciences

Supervisor: Jacques Mehler

SISSA, Trieste - Italy
February, 2012

Acknowledgements

This thesis, as others, has a long story. I left Costa Rica without even knowing where Trieste was, without having any good knowledge about cognition, knowing zero Italian, and with no certainty that I would pass the exam to do a PhD at SISSA. Almost 5 years later, when I am about to finish, other uncertainties arise: Did I make a good job during these years? Was it useful? Where should I go next?

In a short investigation deep into my thoughts I discovered that independently of what I did (or not), I surely learned and received a lot from wonderful people. I believe it is worth offering a line of this thesis to them. I know that almost nobody reads the acknowledgments of a thesis, but I decided to take some time to write this section anyways. After all, this will be the only section that cannot be proven wrong in future investigations.

I remember meeting John Nicholls and Walter Stühmer in Colombia, I (naively) approached them as when you go to a supermarket and ask where to find the things you want: research, infants, memory. What a lucky girl! Not only they knew the "supermarket" for a long while, but they had also administered themselves shops in a very successful way during their lives. What I mostly admired from them (and one of the greatest lessons in life) is that despite their many duties and their well respected fame, their answer was not limited to: "you can find everything in the fourth aisle", they really made sure I would find the right way to reach what I wanted. I am especially grateful to Wal and also to Sissi because of their words, friendship, company, advices and support in every single step of my academic (and not only) life. Thanks for being there for me.

I am very grateful also to Jacques and Marina. I have learned many things from them. Regarding science, of course, but also about many other interesting topics such as politics, international customs, wines, cheeses and all sorts of food! (I guess this is a good time to confess that I have been trying to replicate some of your fantastic recipes, I am not sure I am going to publish the outcomes but I promise to let you know if I get similar results). And Jacques, I want to thank you also for many (unbelievable) stories and anecdotes that you kindly shared with me during these years. They often appeared in the middle of our discussions of results (or my null results). I learned to appreciate the essence of these stories, at the point of thinking of writing a book out of them! Unfortunately I failed, and only got to write the current thesis during these four years. Still, I aim at remembering some of them for the rest of my life.

I ought to thank my former colleagues from the Lab, because most of the techniques, analyses, theoretical questions, etc. that I know, are inherited from their work. Agnes Kovacs, Erika Marchetto, Erno Teglas, Judit Gervain, Juan Manuel Toro and Mohinish Shukla were and continue to be an inspiration for me. I also enjoyed discussing and planning experiments with Jean-Rémy Hochmann and Ricardo Bion with whom I also began to discover Trieste, its *aperitivi*, cinemas, bars, etc.

I need to profoundly thank my current colleagues and friends in the lab. David Gómez, because he has been an excellent collaborator, always ready to help with the analyses of the data, to discuss, to read my writings once, and again, and again, and to hurry in order to meet any *unexpected* deadline. But in fact it would be unfair to say only that about him, because David has been the witness of my frustrations and a constant support inside and outside the office, he has been the friend that I needed. I would like also to thank Amanda Saksida, because she has this special way of seeing life, a way that turns improbable things into possible ones, and problems into solutions. The other nice thing about her is that she spreads this philosophy of life around; in fact I am almost sure this behavior should be a sort of contagious malady! Many thanks, Amanda!

I also thank Laurence White, Andrea Junyent, Alan Langus, Damián Blasi, Hanna Marno, Alissa Ferry, Yamil Dos Santos, and Mikhail Ordin for their friendship and support in the time I was preparing and writing the thesis. I would like to specially thank Mikhail and Tonya Blowers for taking their time to read some sections of this work, and Laurence for the many insightful comments and suggestions of good controls and alternative interpretations of the edges data!

If I were God I would have already booked a place in the heaven for Marijana Sjekloca. But, because I am not, I can just thank her for the kindness with which she would treat me everyday, making me feel at home. I also thank her for the unreserved administrative guidance since the very first moment in which I was accepted as a pre-doc at SISSA, for her patience and for invaluable friendship. Special thanks to Francesca Gandolfo for her great support, continuous help and motivation; thanks a lot also for bearing me while I was writing this thesis and for the free lessons on how to persuade in written Italian. One day I will use properly the simple and the double consonants, I promise it! Thanks to Alessio Isaja for his technical and personal support. My immersion to Macs would have not been the same without his complains, and my work would have not been possible without his “no” every time I asked him to do something. I am kidding, Ale, your help has been precious to me. Your knowledge about all sorts of apparatuses, gadgets, nets, etc., etc., your presence, and availability always resolved my technological frustrations. Yet, I am mostly indebted because you listened and accompanied my frequent trips to melancholic places in my past and my future. Thanks for being there with me.

I would like to express my gratitude to Francesco Macagno, for his indispensable help at the maternity, recruiting and testing the babies. I am grateful to the medical staff in the Obstetrics and Neonatology Departments of the Hospital including Silvia, Dorina, Maria, Lucina, Anna, and Manuela, for their great collaboration and for all I have learned about newborn babies.

This thesis was possible also to the partial financial support received from Ministerio de Ciencia y Tecnología and Consejo Nacional de Investigaciones Científicas y Tecnológicas of Costa Rica to which I am fully thankful.

I am grateful to Diana Bedolla, Jummi Laishram, Gioia Negri, Zeynep Kasap and Alessandra Cifra, who made my life in Trieste enjoyable. For all our aperitivi, cantine aperte, osmizze, dinners, parties, discussions about life and more… I also thank especially Roma Siugzdaite for the great introduction to Italian and Triestinian culture at the time I arrived, for listening to me and give me strong support every time I needed it during these years, and for being one of the most special friends that I have had.

Emily Ngubia Kuria, Gioia Negri and Mounia Drissi were wonderful flat mates and friends, at the point of making me feel them part of my family. Thanks Ngubia also for encouraging me to go beyond my own limits, to dream and to believe.

Quisiera cerrar los agradecimientos, ofreciendo mi más profunda gratitud a mis amigos en Costa Rica y a mis compañeros del Programa de Investigación en Neurociencias de la Universidad de Costa Rica quienes asistieron a mi vida desde lejos a través de constantes correos electrónicos.

También quisiera agradecer a mi abuelita, mis tíos y tías, primos y primas, cuñado y cuñadas quienes con sus oraciones me ayudaron a la distancia. A mi papá y a mi mamá, a mis hermanos y hermana, sobrinos y sobrina, quienes han sido y serán la columna de mi vida. Ellos disfrutaron los logros y las alegrías, y además soportaron pacientemente mis desconsuelos. Agradezco en especial a mi papá, por ser el modelo a seguir como trabajador incansable, por haber disfrutado conmigo (desde la casa) los distintos viajes y por haber siempre creído en mí. Gracias a mi mamá por sus llamadas atentas y sus constantes cuidados. Ella demostró que a pesar de la distancia, una llamada me puede hacer sentir de nuevo en casa, el día y la hora en que lo necesite. Gracias también por enseñarme con el ejemplo que nunca es tarde para plantearse metas, y para lograrlas! A los dos, gracias por comprender mis anhelos y por haberlos apoyado siempre, a pesar de que eso haya implicado estar lejos de ustedes que son las personas que amo.

Gracias Johnny, Yeimy y Luis Diego, por cultivar en mí siempre la alegría y por creer en los sentimientos. Ustedes son personas únicas para mí, con ustedes puedo compartir sensaciones verdaderamente profundas de risas y de llantos, de ilusiones y de verdades. Gracias por el apoyo incondicional durante todas las etapas de mi vida. En todos estos años hubo muchos momentos en los que me hicieron falta, pero el recordar que ustedes estuvieron siempre para mí desde que crecíamos en la finca, me bastó para seguir adelante. Gracias también por hacer el esfuerzo de venir a visitarme (qué sacrificio, verdad?).

Quiero agradecer a Daniela, Luis Andrés, Jose Emilio y Alonso, porque me recuerdan el valor de la juventud, de la inocencia y de la facilidad con la que un corazón lleno de sentimientos puede dejarse fascinar por cosas simples y profundas. Gracias por contar conmigo a pesar de la distancia y a pesar de los años. Cada día, créanme, mi corazón se despierta y piensa en ustedes.

Vorrei finalmente ringraziare Antonio, perché in poco tempo è riuscito ad aiutarmi nel dottorato, pur senza fare niente di scientifico… ne di esoterico (credo)! Grazie tante per la tua compagnia, la pazienza, l’inspirazione, e per tanti sogni costruiti assieme.

Contents

List of Figures

List of Tables

"...and in addition to the Gods you mentioned I must call… upon Mnemosyne. For practically all the most important part of our speech depends upon this goddess..."

(Critias to Hermocrates. Plato, *Critias* 108 d).

Chapter 1

General Introduction

The human brain is an optimal language acquisition device, predisposed to attend to speech from the earliest stages of development. A few days after birth, infants show a preference for listening to speech over other sounds (Vouloumanos & Werker, 2004), and their brain responds differently to linguistic and to non-linguistic sounds in regions that are homologous to those that mediate language processing in the adult's brain (Peña et al., 2003; Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002; Perani et al., 2011). Thanks to the faculty of language any normal human child, given adequate exposure, develops the ability to understand and generate novel utterances and to communicate concepts to other people, without explicit instruction.

As a species we are concerned with language because it is fundamental for learning. Language allows the construction and the transference of our cumulated experiences and abstract knowledge. The greatest achievements of human societies, which include culture, science and art, would not exist without our language faculty.

It is clear that the human mind must be endowed with some computational machinery that makes language acquisition possible in our species. However, not all mechanisms used by language evolved for linguistic reasons. Language relies on brain structures and cognitive functions that are also present in non-human animals and that are not specific to the language domain (e.g., Ullman, 2004; Endress, Nespor, & Mehler, 2009). Memory is one of these general functions.

The ability to remember is a prerequisite for the successful interaction of each species with the environment. Memory assists learning and, as a consequence, it prepares the species for the selection of the best possible behavior in similar future experiences. This important function of memory explains why it is both needed and preserved across different species. Indeed, human and non-human beings share the mechanisms necessary to maintain information at the synaptic level (Bliss & Lømo, 1973; Bliss & Collingridge, 1993; Cooke & Bliss, 2006), and have similar structures to mediate the storage of information in the brain (Squire, 1992; Bunsey & Eichenbaum, 1996).

However, despite sharing the neural mechanisms and structures that sustain memory, each species' system has evolved in order to store in a particularly accurate way the information which is useful for satisfying specific ecological demands. A good illustration of the species-specific function of memory is found in the Costarican leatherback. The leatherback is the largest sea turtle on earth; it can reach up to 700 kilograms of mass and a shell length of 1.7m. Despite the fact that they are extremely ancient animals (the first leatherback turtle bones date back to over 65 million years ago), they have a modern cosmopolitan migratory lifestyle. Leatherbacks migrate thousands of kilometers every year looking for food. Males never leave the water, but females come back to land for a short time (1.5 hours) to lay eggs a maximum of three times in their lives. Intriguingly, in spite of their active migratory behavior, at the time of nesting leatherbacks remember faithfully the way back to the very beach where they were born (Shillinger et al., 2008), even though the little turtles stayed in the beach only a few days before entering in the see.

This leatherbacks' ability exemplifies how a species can profit from specialized memory capacities to accomplish tasks that other species would not be able to achieve. For example, humans are not biologically capable of remembering a path of thousands of kilometers in the sea, instead, humans can effortlessly learn from the sea of words that surround them, and find their way to create sentences and communicate with others since early childhood. Humans can store thousands of speech sounds and the relevant acoustic properties of these sounds because, like any other species, humans have evolved to remember better and more easily the information that enables their learning and survival.

Consequently, the fact that memory is a basic function of the brain does not imply that the human memory for linguistic stimuli is not special. Humans and only humans have the ability to go beyond the acoustic realization of a sound to remember information about it. In fact humans are even able to elaborate mental scenes using the information transmitted through speech. For example, when Cervantes describes his thin, tall and idealistic protagonist, Don Quixote, we benefit from language to visualize this character and to create memories out of this description; there is no need to see Don Quixote or shake his hand to be able to remember him.

The interaction between memory and language processing or language learning has been extensively studied in adults. The first studies on human memory, reported more than a hundred years ago, already focused on the properties of verbal memory (e.g., Ebbinghaus, 1964). Ever since, the acquisition and storage of linguistic sounds continues to be an important topic of investigation. Most research has focused for instance on the short-term memory capacities that sustain the acquisition and retention of word sounds, revealing certain memory mechanisms that might play a crucial role in aspects of phonology, syntax, and word learning.

Even though the various aspects of verbal memory are crucial for understanding the process of language acquisition, there are very few studies addressing the question of how storing word sounds is achieved in developmental populations. The actual capacities and limits of verbal memory in young infants have often been assumed or indirectly inferred from performance. They are frequently mentioned as a possible cause of infants' failures to carry out specific tasks, but only in very few cases have they been the main focus of language development studies (Swain, Zelazo, & Clifton, 1993; Valiante, Barr, Zelazo, Papageorgiou, & Young, 2006).

Research into the neural foundations and properties of newborns and infants' memory was (and probably continues to be) delayed with respect to the adult literature due to conceptual and methodological reasons. In fact, it was assumed for a long time that infants' representation capacities become functional very late in development. Influential theories viewed infants as purely sensorimotor organisms that lack concepts or representations of the world before the age in which they manipulate objects (Piaget, 1952), or produce speech (Sinclair, 1973). In this context, studying representational capacities in preverbal infants was not encouraged. Moreover, the fact that the original methodologies used with adults required participants' explicit retrieval (i.e. pronunciation) of the material, made it difficult to adapt these methodologies to preverbal infants who are unable to talk.

In recent decades many studies have focused on preverbal language abilities. However, psychologists and linguists interested in understanding the first stages of language acquisition have mainly centered their studies on language itself. The majority of work has explored newborns' auditory sensitivities, discrimination capacities, or abilities to extract linguistically relevant properties from the speech signal (e.g., Mehler & Dupoux, 1994; Jusczyk, 2000 for reviews). Neither students of language acquisition nor students of memory development have focused their attention on the specific subfield of memory for speech, despite its importance for understanding our species. This gap not only contrasts with the great number of studies on the development of memory in the visual and motor domains (e.g., Moscovitch, 1984; Oakes & Bauer, 2007; Courage & Cowan, 2009), but also, as mentioned above, with the enormous amount of research focusing on adults' verbal memory (e.g., Crowder & Morton, 1969; Morton, 1969; Morton, Hammersley, & Bekerian, 1985).

The current thesis aims to contribute to filling this gap, arguing that the properties of the representational system in young infants must be explored in order to understand the initial predispositions to learn the most important sound for the species (speech). The thesis attempts to explore features of verbal memory in young infants, in a historical moment in which new methodologies have become available to assess verbal memory capacities in young learners (see Chapter 2) and when a number of studies have suggested that newborns and young infants display abilities in the auditory domain that cannot be conceived without a basic capacity to remember sounds (e.g., Mehler, Bertoncini, Barrière, & Jassik-Gerschenfeld, 1978; DeCasper & Fifer, 1980; Nazzi, Bertoncini, & Mehler, 1998; Mampe, Friederici, Christophe, & Wermke, 2009).

The present work is largely inspired by traditional studies on verbal memory in adults (Ebbinghaus, 1964). It will concentrate on certain features that have been extensively studied in the adult literature (e.g., interference, consolidation, sequence learning), which will be briefly reviewed in the following section. The theoretical framework on verbal memory in adults constitutes a basic reference for the comparison between adult memory and infant memory that will be discussed in the following chapters; it serves to evaluate whether the properties of verbal memory change during a person's lifespan, whether the format and computations are continuous throughout development and to what extent language experience influences the retention of new speech sounds. This is not meant as an extensive review because a more detailed report of the literature will be given in the introduction of each chapter. The general view provided here, together with the empirical work of Chapters 2, 3 and 4, should help to clarify whether models of verbal retention in adults can be applied to young infants.

Some features of verbal memory in adults

The first experimental work on verbal memory and its properties is attributed to the German philosopher Hermann Ebbinghaus (Ebbinghaus, 1964). One of Ebbinghaus' concerns was how to reduce the effects of previous knowledge over new learning. He invented the notion of the nonsense syllable, an attempt to investigate learning in the absence of prior knowledge or associations with the meaning of words. He used himself as the subject for the studies; he read aloud lists of nonsense syllables, and then tried to recite the series from memory. Ebbinghaus' paradigm allowed him to test a multitude of variables that influence verbal memory. For instance, one of his important contributions was to show that items at the beginning and at the end of lists were easier to learn than items in the middle, a phenomenon known as the serial position effect. He also introduced strict controls regarding the timing, the number of study trials, and retention intervals (to study forgetting). Using this methodology, he was able to define the famous forgetting curve, showing very rapid losses over the first few hours or days, and a more gradual decline over the subsequent days and weeks. He also showed that studying the lists at different moments was more effective for long-term retention than massive study episodes. Since then, there has been a huge amount of research in psychology and neurosciences regarding the storage of verbal material. Many studies have improved our understanding of the cognitive and neural mechanisms that underlie verbal learning and memory systems in general. Yet, most of the phenomena that became apparent in Ebbinghaus' investigations have been replicated many times and constitute landmarks in the field.

A number of theoretical models have been proposed over the years to study memory in adults (e.g. Atkinson & Shiffrin, 1968; Crowder & Morton, 1969; Craik & Lockhart, 1972; Baddeley & Hitch, 1974; Morton et al., 1985; Henson, 1998). However, for simplicity, we will define our theoretical standpoint by contrasting it with a single broad framework –the multicomponent working memory model (Baddeley & Hitch, 1974)– that provides an account for how learning novel words is achieved through working memory. Nevertheless, it is important to bear in mind that most –if not all– the studies and theories that gave rise to this and other models were not originally conceived to account for verbal memory in preverbal infants, therefore there may be a subsequent need to postulate a different model for developmental studies.

According to the multicomponent model of Baddeley and Hitch (1974), verbal information is retained in short-term memory by means of an articulatory or phonological loop. This phonological loop has two subcomponents: a short-term phonological store and an articulatory rehearsal process. The memory trace supposedly decays in the phonological store within a few seconds unless the person activates the second component through vocal or subvocal rehearsal. The articulatory component is not only able to refresh the memory trace but can also register visually presented but pronounceable material in the phonological store by means of articulation.

This model accounts for a great number of effects that adult participants experience when trying to maintain information in verbal short-term memory (e.g., the phonological similarity effect by Conrad & Hull, 1964, and the word-length effect by Baddeley et al., 1975) however it cannot be transferred to study verbal learning and short-term memory in preverbal participants; in fact infants, to the best of our knowledge, do not posses the means to rehearse verbal information through articulation because they cannot pronounce words yet. Given that no other model describes in detail the way speech information is maintained in young infants, an attempt will be made in the closure of this thesis to outline such a model.

Operational framework to study verbal memory in preverbal infants

A well-established fact from the adult literature that concerns this thesis is the separation of memories according to the temporal dimension. Early neuropsychological discoveries (Shallice & Warrington, 1970; Baddeley & Warrington, 1970) strengthen the view that memory should not be considered as an unitary system but rather as two separate systems (e.g., Atkinson & Shiffrin, 1968). The short-term memory system is used to retain material over a few seconds or minutes whereas long-term memory involves the retention of information over longer intervals.

Given that this distinction between short-term and long-term memory is widely accepted and frequently used to specify the type of memories under study, we will first try to define the use of these concepts within this thesis. Long-term memory, when mentioned here, will refer to the episodic and semantic information stored and accumulated through linguistic experience and learning. It should be emphasized that unlike infants, adults in most –if not all– situations, are likely to receive top-down influences when performing short-term tasks. Even if presented with meaningless sounds, adult participants come into a given situation prepared with learning strategies and specific associations that they unavoidably use in the laboratory. Considering this situation, and given that infants and newborns possess a much smaller lexical repertoire or verbal experience than adults, it is assumed that long-term memories do not influence newborns' and infants' performance on verbal short-term memory tasks. This assumption implies both that infants do not benefit from the positive effects of linguistic knowledge (e.g., syntax, meaning; Hulme, Maughan, & Brown, 1991; Miller & Selfridge, 1950) and also that negative effects due to previously established semantic or episodic memories (Schacter, 2002) can induce forgetting only in adults. This remark will become particularly important in the subsequent chapters when we compare adults' versus infants' short-term memory capacities.

Short-term memory (which will be the main focus of this work) will be defined here as the information storage system that temporarily retains the traces of a frequently heard speech item (the sound of a word). Recognition is defined as the capacity to identify a previously heard sound (that is supposedly stored in the system), and to distinguish this sound from a newly presented one. An additional feature of short-term memory, as operationally defined in the thesis, concerns its functionality. It is assumed that only the information that newborns and young infants represent in the storage system can be potentially learned. Other speech processing capacities (e.g., phonetic discrimination) are considered a prerequisite for language learning but only when the critical information can be remembered for a few minutes, otherwise the information would be considered unavailable for learning.

The initial state of memory for language in humans: the aim of the thesis

This thesis investigates the origins of verbal memory in humans. It focuses on the interaction between the language faculty and the properties of the representational system in preverbal infants. It proposes empirical evidence to argue that the initial state of the human representational system and the biological disposition to learn language are intimately related at birth. It shows that the newborn's auditory memory favors the storage of information that is crucial for learning language, and vice versa that apparently arbitrary aspects of the language development path may result from general properties of the memory systems and its constraints.

The thesis is organized as follows. Chapter 2 explores the initial state of the human representational system. Previous studies have shown that newborns readily discriminate and extract information from speech (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Dehaene-Lambertz & Peña, 2001; Nazzi et al., 1998; Cheour et al., 2002; Gervain, Macagno, Cogoi, Peña, & Mehler, 2008), and that cortical areas related to speech processing are already functional in the first hours of life (Peña et al., 2003; Telkemeyer et al., 2009). In our studies we ask whether the memory competences of newborns could allow them to remember -not only process or distinguish- detailed information about the words they hear. The ultimate question is whether newborns can actually engage in learning speech from the information they are able to encode, and if so, how this information determines the learning mechanisms that newborns can use to acquire their native language. We address these issues in a series of functional near infrared spectroscopy (fNIRS) studies with neonates, which also allow us to explore whether the brain areas and cognitive mechanisms known to support memory for verbal information in adults could be supporting verbal memory in newborns as well.

In Chapter 3, we investigate the mechanisms that prevent the speech information from being retained in the storage system, leading newborns to forgetting. We seek to answer two main questions. First, how the different sounds from the newborns' surroundings can affect the traces of words in the storage system, and second, which factors can favor or prevent forgetting in newborns. Using fNIRS we show evidence for the existence of a bidirectional interaction between the language faculty and the early memory capacities in humans. On the one hand, certain properties of verbal memory at birth could be human-specific (and applicable only to the language domain). On the other hand, general properties of memory (e.g., consolidation) seem to influence young infants' ability to learn the sound-meaning associations in the first stages of development.

Chapter 4 addresses the issue of whether general properties of memory affect 7-month-olds performance on word-learning tasks. We present a set of eye-tracker experiments assessing classical serial position effects in preverbal infants who begin to learn sound-referent associations. The studies aim at exploring some properties of memory when the phonological categories of the native language have just emerged.

Finally in Chapter 5, we discuss and synthesize the empirical results, proposing an integrative view of verbal memory capacities in the first stages of development; we discuss the broader implications of the findings for the studies of language acquisition and conclude with a preliminary model of short-term representations in preverbal infants, which may serve as a framework for future developmental studies on verbal memory.

Chapter 2

What do newborns remember from words?

Newborns spend most of their time sleeping; when they sporadically wake up they mostly cry or eat. No wonder why until fairly recently, human newborns were thought to be as cognitively incompetent as they were morally innocent. Even William James, one of the greatest psychologists in history, was under the impression that neonates are born in a state that makes them experience the world as *“one great blooming buzzing confusion”*.

In the last decades the view that newborns’ mind is a *tabula rasa* has been challenged. In the domain of language acquisition, for instance, linguists, developmental scientists, cognitive psychologists and neuroscientists have investigated the way infants engage in acquiring language, showing the core dispositions of the human species to attend to speech, and the functional specialization of certain areas of the brain to process linguistic stimuli since birth (e.g., Mehler & Dupoux, 1994; Jusczyk, 2000 for reviews). The present chapter follows this approach by investigating the basic representational abilities of human newborns. The studies explore whether the competences of newborns could allow them to remember -not only process or distinguish- detailed information about the words they hear.

These early cognitive abilities should provide insights into the toolbox with which the human mind comes equipped to start learning about the environment. Evaluating auditory memory in newborns should also contribute to determining the amount and kind of information that the human learner can represent when they begin acquiring the native language.

A number of studies have shown that newborns and fetuses recognize some general acoustic properties of speech. For example, few minutes after birth, newborns prefer a story heard during the last weeks of pregnancy to a new story (DeCasper & Spence, 1986), they also prefer their native language to a foreign language (Moon, Cooper, & Fifer, 1993), and their mother’s voice to a stranger’s voice (Mehler, Bertoncini, Barrière, & Jassik-Gerschenfeld, 1978; DeCasper & Fifer, 1980). However, it is unlikely that fetuses remember specific elements (i.e. sound form of words), because the properties of the uterus render many phonemic differences imperceptible, allowing only prosodic information to reach the fetus (Querleu, Renard, Versyp, Paris-Delrue, & Vervoort, 1988). The fetus, only hearing a series of low-passed-filtered sounds, is at birth all of a sudden overwhelmed with more complex stimuli containing detailed phonetic information. For the first time, the newborn brain has access to enough acoustic detail to distinguish all words in the surrounding language. Can newborns remember enough information to distinguish among different words?

Indications of the newborns’ capacity to remember speech stimuli come from a pioneer behavioral study in which memory for words has been tested. Swain and colleagues (Swain et al., 1993) explored that ability of 1-2 day-olds to remember a word over 2 consecutive days. The authors found that newborns that were familiarized with the word responded less to this word on the second day compared to a control group of infants that heard the word for the first time on the second day. However, newborns that heard the familiar word were exposed to the same word for up to 32.5 minutes (975 repetitions of the word). It is therefore unknown whether neonates can encode a word in shorter and more ecological exposures. Moreover, from this study it was difficult to infer the kind of information participants relied on to distinguish the familiarization word (e.g. *tinder*) from the test word (e.g. *beagle*), since these sounds differ from each other in many acoustic and phonetic features.

In this chapter we will introduce a fNIRS paradigm designed to assess auditory memory in neonates (as described in Benavides-Varela, Gómez, & Mehler, 2011). Using this paradigm we will explore the basic word recognition capacities and the associated cortical responses in the newborn' brain (Experiment 1). Moreover, an investigation of the specific information that newborns encode from a word will follow in Experiment 2. A number of aspects differentiate this study from previous investigations on newborn's memory for words. We used fNIRS, a technique suitable to study how neonates process auditory stimuli (see next section). We also ensured that familiarization and test words were matched to have the same acoustic properties. Moreover we implemented a much shorter familiarization phase than previous studies (six minutes in total, including more than four minutes of silent pauses).

The fNIRS technique and its advantages for testing memory in young infants

In the last fifty years, behavioral studies yielded important findings on relevant topics for language acquisition such as infants' discrimination capacities, categorization of speech sounds, recognition of properties of their native language, etc. (see Jusczyk, 2000, for a review). Even though methods relying on infants' behaviors (e.g., head orientation, looking time, sucking rates) have been crucial for understanding preverbal cognition, they were not suitable for investigating topics such as attention or memory. We will try to illustrate why these methods were not adequate to test memory using the discrimination studies as an example.

In a standard discrimination study, each participant is exposed to a repeated presentation of a sound. As the stimulus becomes more and more familiar, the infant's response to the stimulus (looking, orienting or sucking) decreases. The same stimulus continues to be presented until the infant reaches a certain habituation criterion (usually a decrement of about 25% of the initial response). Immediately after achieving this criterion, infants are presented with the same stimulus, to which they are expected to respond less, or with a novel stimulus to which they are expected to show an increment of the response –provided that they are able to discriminate between the familiar and the novel sounds–.

As opposed to discrimination studies, in which the test item immediately follows the habituation phase, in memory studies the habituation and the test phases need to be separated by a pause. Thus, by using a design similar to the one described above, one could theoretically turn a discrimination study into a memory study simply by including a pause between habituation and test. However, many attempts to try to apply this modification failed in the past because infants often fell asleep during the silent pauses, or because the behavioral responses of the control groups became noisy (Jusczyk, Jusczyk, Kennedy, Schomberg, & Koenig, 1995). The necessity of a criterion to shift from the habituation to the test phase constituted an additional limitation of the behavioral methods to study some variables that influence memory, for example the amount of exposure. Indeed, the amount of experience with the sound should influence the robustness of the memory trace (the longer the exposure, the solider the memory), therefore a design in which all participants hear a sound the same number of times becomes desirable. The habituation-dishabituation method was not appropriate to test the effects of exposure over memory because the amount of stimulus presentations differed considerably across infants (each participant required a different number of repetitions in order to achieve the habituation criterion).

Neuroimaging methods constitute an alternative for studying memory in young infants, offering a number of advantages with respect to behavioral methods. First, they measure the neural responses without requiring any overt behavior from the infant. This property facilitates the observation of abilities that might have been undetected in the past or that were very difficult to explore in previous behavioral investigations. Moreover they allow us to test infants who have even fallen asleep, for instance during the silent pause. Second, because there is no need to have a habituation criterion, in the neuroimaging studies all participants arrive to the test having heard the familiarization sound the same number of times. In addition, neuroimaging methods offer the possibility to investigate aspects that influence memory such as time of study and length of the retention interval by shortening and lengthening the familiarization and the pause across studies. A further advantage is the lower rejection rate in the neuroimaging studies than in behavioral studies. Finally, brain-imaging techniques not only provide a yes/no answer to the experimental question at hand, but also inform researchers of the activation of several brain areas, providing a link between behavioral and neural responses.

Despite these advantages, few neuroimaging techniques are fully appropriate to work with healthy newborns or young infants. This is because the experimental equipment produces high levels of acoustic noise, because they require the application of liquids or gel in the newborn's head, or because of the very low tolerance to facial and body movement. Functional near-infrared spectroscopy (fNIRS) is a very suitable technique to study language faculties and cognitive capacities in the newborn's brain (see Aslin & Mehler, 2005; Minagawa-Kawai, Mori, Hebden, & Dupoux, 2008; Lloyd-Fox, Blasi, & Elwell, 2010; Obrig, Rossi, Telkemeyer, & Wartenburger, 2010; Gervain et al., 2011 for reviews on fNIRS developmental studies). This technique is fully non-invasive, it does not require the use of carrier substances to keep the device in place on the infant's head, and there is no need to use radioactive substances or expose infants to magnetic fields. fNIRS has the disadvantage that the localization of specific areas remains somewhat imprecise due to its low spatial resolution, and in our experiments also due to the lack of structural imaging to guide the probe placement. On the other hand the device makes hardly any noise, which makes it handy for testing young participants with auditory stimuli.

In the following section, we will begin our exploration of newborn's memory abilities by borrowing habituation-dishabituation designs from the infants' research tradition, and by adapting them for the use with fNIRS. The technique works by measuring changes in cerebral blood flow volume and oxygen saturation using optical means. It is based on the emission of near-infrared laser light on the subject's scalp and its absorption by the blood chromophores: oxyhemoglobin (oxy-Hb) and deoxyhemoglobin (deoxy-Hb) (Meek, 2002; Villringer & Chance, 1997). The changes in concentration of the chromophores provide a means of investigating brain function. The rationale is that during neural activation the metabolic demands change and provoke an increase in oxygen consumption, local cerebral blood flow (CBF) and oxygen delivery. These hemodynamic changes in the brain can be then detected with functional near infrared spectroscopy.

2.1 Experiment 1: Newborn's brain signals memory for words

This first experiment has the purpose of establishing whether newborns can remember a previously heard word and can distinguish it from another acoustically similar word after a silent pause. Additionally, we attempt to explore the cortical brain structures that are involved in the recognition of speech sounds at birth. For these purposes we will use fNIRS to test newborns in a simplified recognition memory task. We ask whether exposure to familiar and novel words lead to different hemodynamic responses in the cortical areas of the brain.

2.1.1 Participants

Fifty-six neonates (27 females, mean age 3.1 days, range 1-5 days) were included in the analysis. Thirteen additional neonates were tested but excluded from the analysis because of head movements that produced large motion artifacts (n=7), or because they cried before the end of the experiment (n=6). Neonates were recruited at the newborn nursery at the Azienda Ospedaliera Universitaria Santa Maria della Misericordia in Udine, Italy. Neonates were considered eligible if they had gestational ages between 38 and 42 weeks, Apgar scores ≥8 in the first and the fifth minutes, diameter of head ≥33.5cm and normal history during pregnancy and delivery. We failed to obtain signals from neonates who had thick hair or who had cephalhematoma, therefore these became exclusion criteria in this and the following studies. Bioethics Committee of SISSA/ISAS (International School for Advanced Studies) approved the study. All parents signed an informed consent before the experiments.

2.1.2 Stimuli

Two pseudo words (*mita, pelu*) were recorded from a female native speaker of Italian. They were both pronounced using child directed speech (*mita*: mean pitch 281Hz, pitch range 191-478Hz; *pelu*: mean pitch 253Hz, pitch range 190-490Hz), they carried first syllable stress, had a CVCV structure, and were edited to have the same mean intensity (70dB) and duration (700ms).

2.1.3 Procedure

The experiment consisted of a familiarization phase, a silent interval, and a test phase (see figure 2.1). The familiarization phase lasted 6 minutes and was organized in 10 blocks. Each block contained 6 identical words. Within blocks, words were separated by pauses of randomized length (0.5s or 1.5s), yielding blocks of approximately 10s each. Blocks were separated by time intervals of varying duration (25s or 35s) to avoid synchronization between stimuli occurrences and spontaneous oscillations. A 2-minute silent interval was inserted between the end of the familiarization and the beginning of the test phase. The test lasted for 3 minutes and it consisted of five blocks. In the test, half of the infants heard the same word that they heard in the familiarization, whereas the other half was presented with a novel token. The words used in the familiarization were counterbalanced across participants. For half the infants, the stimulus of the familiarization phase was *mita* and the novel word in the test was *pelu*. For the other half of participants, the words were exchanged, so that *pelu* was used during the familiarization phase and *mita* was presented as the novel word in the test. The statistical analyses used pooled data from newborns that were familiarized with both *mita* and *pelu*, since there was no significant difference between participants in any of the familiarization blocks (permutation test, all *ps* > 0.30), evidencing that the acoustic properties of the words *per se* are not responsible for different neural responses.

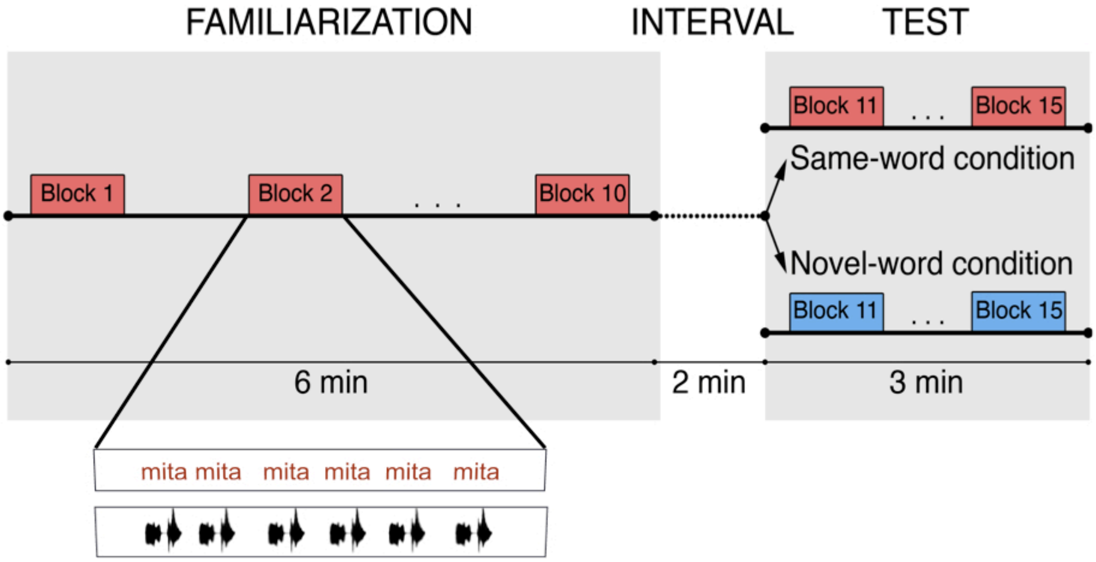

Figure 2.1. Schematic diagram of the procedure used in Experiment 1.
During the familiarization phase, all the neonates were presented with 10 blocks composed of 6 identical words. In the test, 5 blocks of the familiarization word were presented to half of the neonates whereas the other half listened to a novel word.

Neonates were tested while lying in their cribs, asleep or in a state of quiet rest. A nurse or a medical doctor assisted neonates inside a dimly lit sound-attenuated booth where the experiment was run. Sound stimuli were presented via two loudspeakers placed at a distance of 1.2m from the infant's head at a 30° angle on both sides, raised to the same height as the crib. The speakers were connected to a Macintosh power PC G5 computer that at the same time operated the NIRS machine and presented the auditory stimuli using PsyScope X software (http://psy.ck.sissa.it/). Both the NIRS machine and the computer were placed outside the experimental booth and were controlled by the experimenter. An infrared video camera was used to monitor the infant's behavior.

2.1.4 Data acquisition

A NIRS machine (ETG-4000, Hitachi Medical Corporation, Tokyo, Japan) was used. The total laser power output per fiber was 0.75mW. A silicon holder (probe) kept the optical fibers in place. In each probe, five fibers were emitters and four were detectors, arranged in such a way that allowed simultaneous recording from 12 points (channels). Two probes were used, one for each hemisphere, giving a total of 24 channels. The separation between emitters and detectors was 3cm and the sampling rate 10Hz. Probes were placed on the neonate's head by using skull landmarks. Although individual variation cannot be excluded, placement maximizes the likelihood of monitoring the temporal, parietal and frontal areas. Channels from 1 to 12 were placed on the left hemisphere and from 13 to 24 on the right hemisphere. Channels 1, 2, 4 and 5 were roughly located in the left frontal regions; 3, 6, 8 and 11 in the left temporal and 7, 9, 10 and 12 in the left parietal region. Channels 13, 14, 15 and 16 were in the right frontal area; Channels 17, 19, 22 and 24 in the right temporal and 18, 20, 21 and 23 in the right parietal region (see figure 2.2)

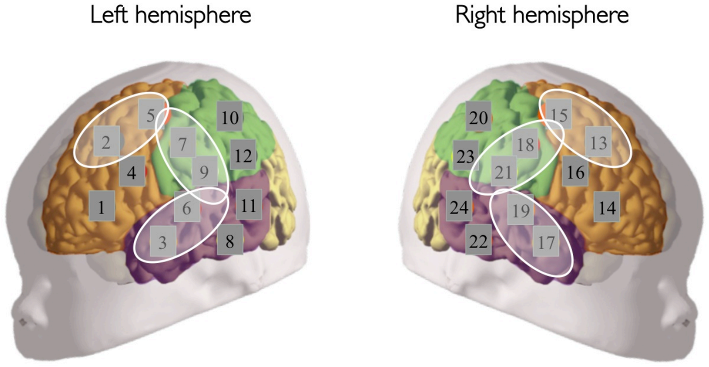

Figure 2.2 Location of channels (grey squares) and regions of interest (white ellipses) on a schematic neonate brain

2.1.5 Data processing and analysis

Our analysis was based on the variations in oxy-hemoglobin concentrations assessed on the basis of the light absorption recorded by the NIRS machine. The signal was band-pass filtered between 0.02Hz and 1.00Hz to remove components arising for instance from slow fluctuations of cerebral blood flow, heartbeat (normally 70-190 beats per minute: about 1.1-3.1Hz) and other possible artifacts (see also Peña et al., 2003). Single blocks from specific channels were eliminated on the basis of two criteria: 1) light absorption of less than 1% of the total light emitted (generally because the fibers were not touching the neonate's scalp); 2) the presence of large movement artifacts. The criterion to detect artifacts was the presence of rapid changes in the signal (> 0.1 mmol*mm in an interval of 0.2s). Blocks with more than 12 rejected channels were excluded[1]. Participants were included in the analysis only if the amount of data rejected was less than 30%.

For the non-rejected blocks, a baseline trend was linearly fitted between the mean of the 5s preceding the onset of the block and the mean of the 5s between the 25th s and the 30th s after the onset of the block. For each channel, the mean signal changes in the period between the end of the auditory stimulation and the following 9s were used to carry out the subsequent statistical analysis. This analysis period corresponds to the time window in which the maximum amplitudes of the hemodynamic responses are expected based on previous studies (e.g., Peña et al., 2003; Kotilahti et al., 2005; Taga & Asakawa, 2007).

---

[1] Although this rejection criterion did not discriminate the location of the channels, we generally observed that the number of channels included in the analysis was equally distributed in the two hemispheres.

To assess whether the two conditions differed across the familiarization phase, we computed the maximum difference between the mean activation for each condition. That is to say, for each block b=1,2,…, we calculated

$$\mathrm{Diff}(b) = \max_{c=1,\ldots,24} \left| \overline{A}_1(b,c) - \overline{A}_2(b,c) \right|$$

where $\overline{A}_j(b,c)$ is the activation in block b and channel c, averaged among all neonates assigned to condition j (Same-word/Novel-word). When analyzing the familiarization phase as a whole, we further computed Diff(fam) as the maximum of Diff(b) for all blocks b=1,2,…,10. Because the distribution of this statistic is not Gaussian, we evaluated significance using non-parametric methods. Specifically, we used permutation tests (Nichols & Holmes, 2002). In these tests, a distribution for the test statistic under the null hypothesis is obtained by re-randomizing the condition assigned to each subject. Assuming that the two conditions do not differ during the familiarization phase, then the distribution of Diff(fam) is the same for the original group assignment as for any random reassignment. Significance is then computed as the proportion of reassignments exhibiting a value of Diff(fam) greater or equal than the one associated to the original groups. Other statistics such as Diff(test) are built and statistically evaluated in the same way. We used 10,000 random reassignments for each permutation test.

Additional tests were conducted to identify the channels contributing to the differences in the test phase that the previous analysis found. We compared the two groups on a channel-by-channel basis using 2-sampled t-tests. To solve the problem of multiple comparisons, we computed corrected p-values based on the procedure proposed in (Benjamini & Yekutieli, 2001; Theorem 1.3), to control the False Discovery Rate at the 5% level. That is, starting from the 24 uncorrected p-values (one per channel) $p_{(1)}, p_{(2)}, \ldots, p_{(24)}$ sorted from smallest to largest, corrected p-values were obtained as $p_{(k)}^{FDR} = \frac{24 \cdot c(24)}{k} p_{(k)}$, with c(24) defined as 1 + 1/2 + 1/3 + … + 1/24.

To examine brain areas associated with auditory memory, we compared activation during the last block of the familiarization phase in frontal, temporal and parietal areas. We used a repeated-measures analysis of variance (ANOVA) with Condition (Same-word/Novel-word) as a between-subject factor and Area (frontal/temporal/parietal) and Hemisphere (left/right) as within-subject factors. Two channels were included in each area-hemisphere region. In the left frontal, channels 2 and 5; left temporal, 3 and 6; left parietal, 7 and 9; right frontal, 13 and 15; right temporal, 17 and 19; and right parietal, 18 and 21 (see figure 2.2). These channels were chosen based on previous imaging studies of auditory processing in neonates and young infants (Gervain, Macagno, Cogoi, Peña, & Mehler, 2008).

2.1.6 Results

Significant differences in brain activation between the Same-word and Novel-word conditions were observed in the first block of the test (permutation test, $p<0.0001$ for oxy-Hb ; $p<0.01$ for deoxy-Hb). This was the only block in which significant differences in brain activity between neonates in the Novel-word and in the Same-word conditions were found in the test. There were no significant differences between the conditions in the familiarization phase. In addition, differences in the responses from the first block of the test and the last block of the familiarization showed a decrement of oxyHb in the Same-word condition and an increment in the Novel-word condition (permutation test, $p < 0.01$) (Figure 2.3).

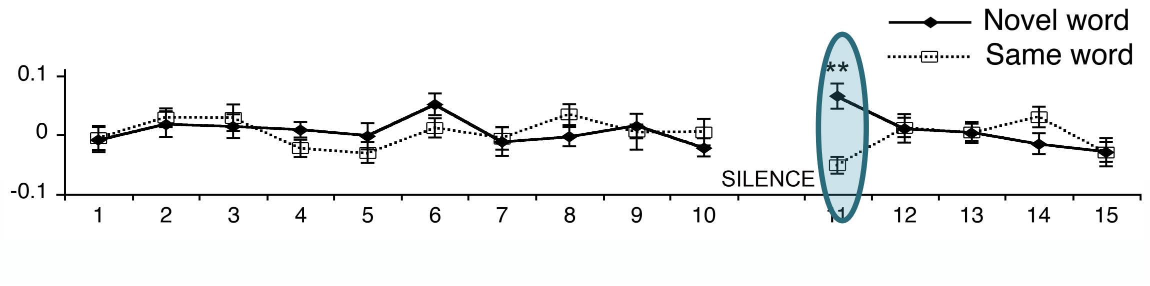

Figure 2.3 Time courses of the relative hemodynamic changes in Experiment 1 (oxy-Hb). There significant differences between the conditions in the first block of the test phase. Data is averaged across all the channels and subjects per group. The dashed line indicates the time series for the group that heard the same word before and after the pause; the continued line represents the group that heard a novel word in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-Hb in mmol*mm is displayed. **, p <0.0001.

For deoxy-Hb, we also observed significant differences in the responses from the familiarization to the test (p<0.05), but the magnitude of the changes in concentration was smaller as compared to oxy-Hb and the pattern of activity was inverted. There was an increment of deoxy-Hb in the Same-word condition and a decrement in the Novel-word condition.

In order to determine the channels and brain areas that contributed to the overall difference found in the test phase, we compared the activation elicited by the two conditions during the first test block in each of the 24 channels. The data show that word recognition evokes a diffuse cortical response in the neonate brain, which is bilaterally spread over temporal, parietal, and frontal areas. Given that fNIRS is not suitable to measure changes in deeper brain areas, it remains to be established whether the observed pattern is only part of a larger network in the newborn brain. Moreover, it should be clarified whether this activation is solely responsible for word recognition.

In addition, we compared brain responses –as measured by oxyHb changes– in six regions of interest: frontal, temporal, and parietal regions of the left and right hemispheres (Figure 2.4). We observed a main effect of Condition [ANOVA, F(1,54)=18.469; *P* = 0.0001] due to a greater brain activation of the neonates in the Novel-word condition as compared to neonates in the Same-word condition. There were no main effects of Hemisphere [ANOVA, F(1,54)=1.231; n.s.] or Area [ANOVA, F(2,108)=1.154; n.s.], and no significant interactions between factors [all *Fs* < 1].

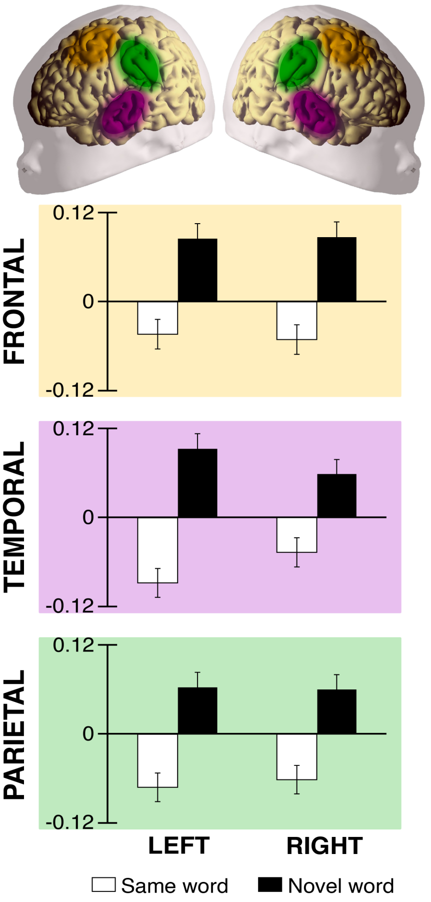

Figure 2.4 Experiment 1: Oxy-Hb responses in the first block of the test phase minus responses in the last block of familiarization.
There was a general a decrease in the concentration of oxy-Hb when neonates heard the same word before and after the pause (white bar). In contrast, when neonates were confronted with a novel word in the test (black bar) the concentration from the familiarization phase to the test increased. Colored ellipses on the schematic neonate brain indicate the localization of the channels included in the areas of interest.

2.1.7 Discussion

The ability to recognize the sound of a previously heard word is a necessary prerequisite to learn language. This study shows that newborns possess the means for storing the sound pattern of a word for few minutes. Besides directly supporting vocabulary acquisition, memory for word-sounds may also facilitate learning of the properties of the surrounding language. For instance, in order to discover phonological categories, phonotactic constraints, and other statistical regularities of the signal, a substantial database of speech sounds should be stored.

While previous studies in newborns' memory relied on behavioral responses, we looked at brain hemodynamic activity instead. The fNIRS data support and extend previous findings in several ways. First, the results show that the newborn brain is able to encode a word during a brief exposure. Exposure for over half an hour -or over two hundred repetitions of the words- was used in previous studies looking at memory in newborns (Swain et al., 1993; Valiante et al., 2006). The familiarization phase of this experiment was much shorter (six minutes in total, including more than four minutes of silent pauses), and with fewer instances of the familiarization word, providing evidence that newborns do not require protracted experience to remember a word.

The data also provided a neural signature for word recognition in newborns. This consisted in a decrease of oxy-hemoglobin when the familiarization word was presented again in the test phase, and an increase in the oxy-hemoglobin when newborns listened to a novel word. In the conditions of this experiment, word recognition evoked a diffuse response in the neonate brain, which was bilaterally spread over temporal, parietal, and frontal areas. It is not clear from these results whether this pattern of activation solely reflects memory processes, or whether it also indicates the participation of other cognitive mechanisms related to language processing or attention. Yet, the differential response observed when newborns listened to familiar as opposed to novel words provides an opportunity for studying in depth the neonate's word recognition capacities in the following studies.

Finally, this experiment provides some first insights about the information newborns store when they hear words. Neonates maintain the representation of a CVCV word (or parts of it) over the silent retention interval and contrast this representation with the sound of a new CVCV word presented during test. Importantly, the two words were pronounced by the same speaker and had similar acoustic properties, suggesting that the information encoded by newborns in the storage system goes beyond low-level perceptual features such as pitch, voice quality, duration, or intensity.

The question arises as to the level of detail that the newborns represented from the sound of a word. In the current study, the familiar and the novel words were composed by entirely different sets of phonemes, which makes impossible to evaluate whether newborns represented all the segments of the word, whether they retained a more general/holistic representation of the word, or whether some of the sounds that constituted the word were better encoded than others.

In the following experiment we will explore the basis of word representation in the newborn brain. We will focus on the elements that participants remember from the sound of words. This information may provide a better understanding of the properties of the information storage system at birth, whether they reflect the interaction between memory and language, and in which way memory could impose some constraints in the first stages of language acquisition.

2.2 Experiment 2: The information encoded at birth

Judging by the great number of studies showing newborns' abilities to discriminate fine phonetic contrasts, one might be inclined to ascribe to newborns a very detailed representation of the sound of words. Newborns discriminate nearly all phonetic contrasts of the languages of the world, including those that their parents can no longer distinguish (Eimas et al., 1971). They distinguish consonants differing in place of articulation, voicing, manner of articulation (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Eimas & Miller, 1980; Dehaene-Lambertz & Peña, 2001), duration (Kushnerenko et al., 2001), as well as vowel quality contrasts (Bertoncini, Bijeljac-Babic, Blumstein, & Mehler, 1987; Cheour-Luhtanen et al., 1995). Could these extraordinary discrimination capacities imply that newborns can store all the segmental details of a word? Different studies suggest that, at least in adults and older infants, this is not the case. As it will be briefly reviewed in the following section, despite the auditory discrimination capacities for vocalic and consonantal contrasts, adults and older infants privilege consonants when processing lexical items. Is it possible that a similar bias exist in newborns? Or rather could maturational changes modify the dimensions of speech stimuli that humans encode when processing the sound of words?

The present study aims at uncovering the nature of the newborn's word representations. Knowing the specific information encoded in the newborn's brain should inform us about the initial state of our verbal memory capacities, and the properties and constraints that guide language acquisition across development. Moreover, differences and similarities between newborns' and previously reported adults' language/memory mechanisms could shed light on the learning abilities available for learning language at different ages.

Word processing in older infants and adults predicts consonantal bias

In agreement with the CV hypothesis that proposes two different roles for consonants and vowels in language processing (Nespor, Peña, & Mehler, 2003), a great number of studies suggest that adults privilege consonants over vowels when processing lexical items. Behavioral investigations have shown, for instance, that adults rely more on consonants than vowels for word segmentation tasks (Bonatti, Peña, Nespor, & Mehler, 2005; Toro, Shukla, Nespor, & Endress, 2008; Toro, Nespor, Mehler, & Bonatti, 2008), visual-word processing (Lee, Rayner, & Pollatsek, 2001, 2002; New, Araújo, & Nazzi, 2008), and auditory lexical access (Van Ooijen, 1996; Cutler, Sebastián-Gallés, Soler-Vilageliu, & Van Ooijen, 2000). Electrophysiological and functional neuroimaging studies have also reported different brain activation patterns for consonants and vowels on healthy adult participants (Carreiras & Price, 2008; Carreiras, Vergara, & Perea, 2009). Moreover contrastive patterns of errors in consonants and vowels have been described in neuropsychological studies in patients pronouncing, spelling, or writing words (Caramazza, Chialant, Capasso, & Miceli, 2000; Cotelli, Abutalebi, Zorzi, & Cappa, 2003; Miceli, Capasso, Benvegnu, & Caramazza, 2004), supporting the view that consonants are more important for lexical processing than vowels (Nespor et al., 2003).

The predominant role of consonants for word processing and word acquisition is also observed in older infants. For example, Nazzi and collaborators have shown that 16-month-olds and 20-months old infants could learn simultaneously two pseudowords that differ minimally in their consonants, but failed when the words differ only in their vowels (Nazzi, 2005; Nazzi & New, 2007; Havy & Nazzi, 2009; Nazzi & Bertoncini, 2009). The same bias has been observed crosslinguistically (Nazzi, Floccia, Moquet, & Butler, 2009); and more recently it has also been reported in 12-month-old infants performing word-learning tasks (Hochmann, Benavides-Varela, Nespor, & Mehler, 2011). In sum, evidence from adults and infants indicates that consonant sequences are more important than vowel sequences for the representation of a familiar word.

What about newborns? There are reasons to predict that newborns, in contrast to adults and older infants, could rely more on vocalic than on consonantal information to recognize the sound of a familiar word. As proposed by Mehler and colleagues, newborns may retain only the information that is more salient in the speech signal (Mehler, Dehaene-Lambertz, Dupoux, & Nazzi, 1996). Vowels are longer and louder than consonants, and they are a rich source of information about the prosodic structure of the language. Encoding vowels –the main carriers of prosody– could be helpful in acquiring not only the phonological but also the syntactic structure of the native language (Morgan & Demuth, 1996; Marina Nespor, Guasti, & Christophe, 1996; Christophe, Nespor, Teresa Guasti, & Van Ooyen, 2003). A predisposition to attend to such carriers of prosody in mother's speech could lead newborns to engage in the interaction with the mother (Fernald & Simon, 1984) which also contributes to the infants' linguistic development (Kemler Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989). Prosodic cues also play an important role in identifying indexical information such as speakers' identity (e.g. the mother's voice DeCasper & Fifer, 1980; Mehler et al., 1978). Furthermore, it is well known that in the course of development infants learn native vocalic categories earlier than consonantal categories (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984; Polka & Werker, 1994; Best & McRoberts, 2003; Kuhl, 2004), suggesting that vocalic information might be available to infants earlier than consonantal information.

This experiment has the purpose of establishing whether newborns can retain detailed phonetic information about a word, and thus whether they can distinguish this word from a phonetically similar one. We constructed the test items such that they differed from the familiar word either in the vocalic-tier or in the consonantal-tier.

This manipulation may allow us to tell apart three different hypothesis derived from previous literature: 1) newborns remember all the phonetic distinctions that they are capable of discriminating, 2) newborns, like older infants and adults, represent better the consonants of the word, 3) newborns encode better the information carried by vowels because this information is more salient and has more energy than the information carried by consonants.

2.2.1 Participants

Forty-four healthy full-term neonates (17 males and 27 females; mean age 2.5 days; range 1-4 days; birth weights 2764-4302g) participated in the study. Seven additional neonates were tested but excluded from the analysis either because their head movements produced large motion artifacts (n=2), or because they cried before the end of the experiment (n=4), or because of an experimental error (n=1). Selection criteria included gestational age between 38 and 42 weeks, Apgar scores ≥8 in the first and fifth minutes, diameter of head ≥33.5cm, absence of cephalhematoma, and think hair. Neonates were recruited from the newborn nursery at Azienda Ospedaliera Universitaria Santa Maria della Misericordia in Udine, Italy. Parents of the infants signed the informed consent after they had understood the procedure and after all their questions had been answered. The Ethics Committee of the International School for Advanced Studies approved the study.

2.2.2 Stimuli

Two CVCV words (*mimi* and *lili*) were used during the familiarization (see next section); the test words (*sisi*, *mama*; *titi*, *lala*), were chosen because they have phonemes that differ from those of the familiarization words in the maximum number of phonetic features (maximum vocalic distance possible in Italian = 3 features from /i/ to /a/; maximum consonantal distance possible in Italian = 4 features, for instance from /m/ to /s/ and from /l/ to /t/). Given that in this study there were subtler differences between the familiarization and the test words than in the previous study, we used words with two repeated syllables. This modification was meant to favor the saliency of the auditory stimuli for the newborns (see Gervain et al., 2008). All words had no meaning in Italian, the language spoken by the parents of the infants. A female speaker recorded the stimuli using a child directed speech. Each word carried stress on the first syllable and was edited to 70dB mean intensity and 700ms duration (see figure 2.5).

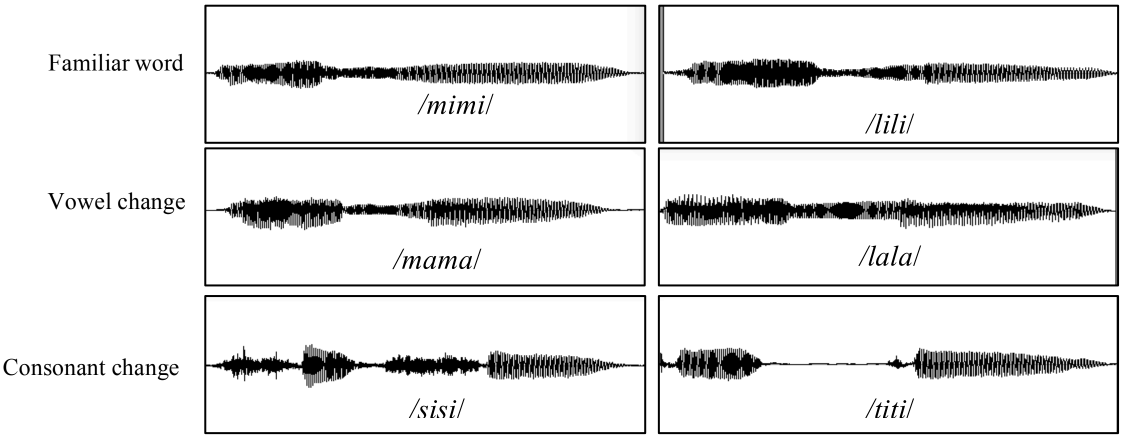

Figure 2.5. Sound waves of the words used in Experiment 2.

2.2.3 Procedure

The design consisted of a familiarization phase of 10 blocks, a silent interval, and a test phase of 5 blocks (Figure 2.6). Each block was about 10s long and contained 6 identical words separated by pauses of randomized length (0.5s or 1.5s). Blocks were spaced at time intervals of varying duration (25s or 35s). The experiment lasted about 11 minutes (6 minutes familiarization phase, 2 minutes silent interval, 3 minutes test phase).

Two words were used during the familiarization phase; half of the participants heard one word (e.g. *mimi*), and half heard the other word (e.g. *lili*). In the test phase, half of the newborns heard a word that preserved the consonants of the familiar word but had a different vocalic tier (Vowel-change Group), whereas the other half heard a word that changed in the consonantal tier, but preserved the vowels of the familiar word (Consonant-change Group). For instance, if the familiarization word was *mimi*, the test words were *mama* for the Vowel-change Group and *sisi* for the Consonant-change Group. Similarly, if the familiarization word was *lili*, the test words were *lala* for the Vowel-change Group and *titi* for the Consonant-change Group.

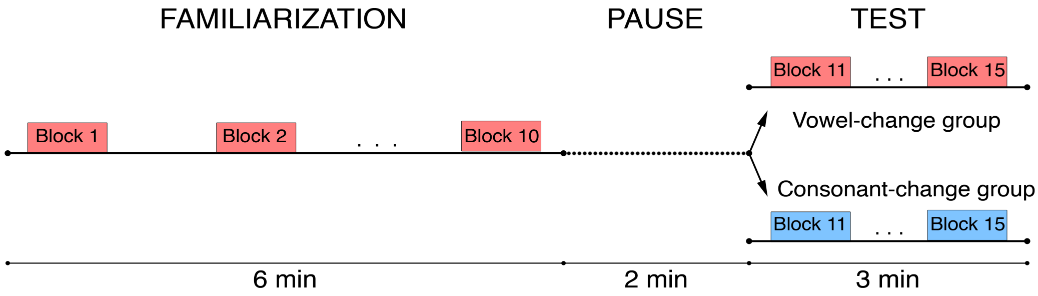

Figure 2.6 Schematic diagram of the procedure used in the Experiment 2.
Each rectangle represents a block composed of a series of 6 words separated by short pauses of 0.5 or 1.5s. Consecutive blocks were separated by silent intervals of 25s or 35s. The experiment consisted in a familiarization phase in which all neonates listened to 10 blocks of stimulation. After a silent retention interval –pause– of 2 minutes, a test phase followed. In the test, half of the neonates listened to 5 stimuli blocks of a word that preserved the consonants of the familiar word but had a different vocalic tier (Vowel-change Group), whereas the other half listened to a word that changed in the consonantal tier, but preserved the vowels of the familiar word (Consonant-change Group).

2.2.4 Data acquisition

The placements of the probes and data acquisition were identical to those used in the previous experiment.

2.2.5 Data processing and analysis

The analyses performed were identical to those conducted for the previous experiment.

2.2.6 Results

We found no significant differences between newborns familiarized with the words *lili* or *mimi* (permutation test, $p = 0.35$), showing that brain responses did not differ as a function of phonological differences of the words used in the familiarization. Data from all the participants were then pooled in the analysis calculated over all 24 channels. The results showed no significant difference between the Consonantal-change and Vowel-change groups in either the familiarization or the test phase (permutation test, $ps > 0.05$) (See figure 2.7).

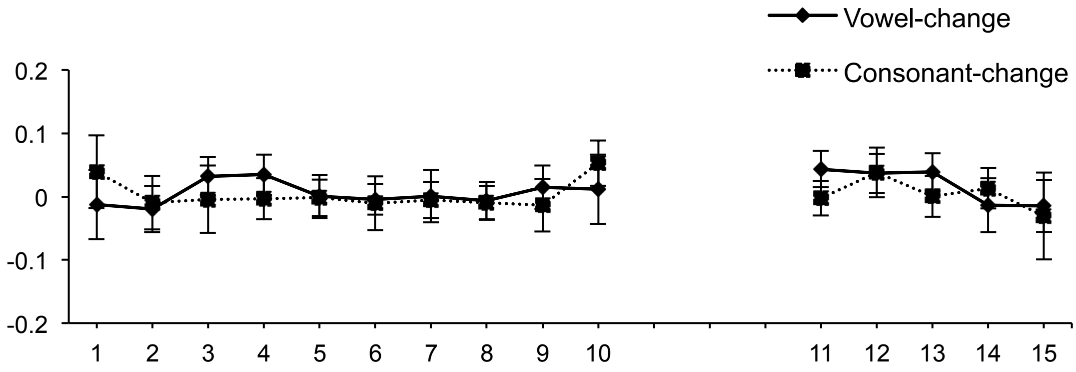

Figure 2.7 Time courses of the relative hemodynamic changes (oxy-Hb) in Experiment 2. ***Data were averaged across all the channels and subjects per group.*** There were no significant differences between the groups. The dashed line depicts the results for the Consonant-change group; the continued line represents the Vowel-change group. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed.

In addition and similar to the previous study, we performed planned comparisons in the first block of the test phase in channels grouped in six areas of interest (see section 2.1.5). The specific permutation tests performed in these regions showed a significant difference between the groups in the right frontal area. Neonates in the Vowel-change group displayed a greater response than neonates of the Consonantal-change group (permutation test, $p < 0.001$) (See Figure. 2.8). Similar results were found in the right frontal area for the separate comparison between newborns that were presented in the test with *mama* or *sisi* (permutation test, p <0.01); as well as newborns that heard *lala* or *titi* (permutation test, p <0.05). No other area showed significant differences between the groups.

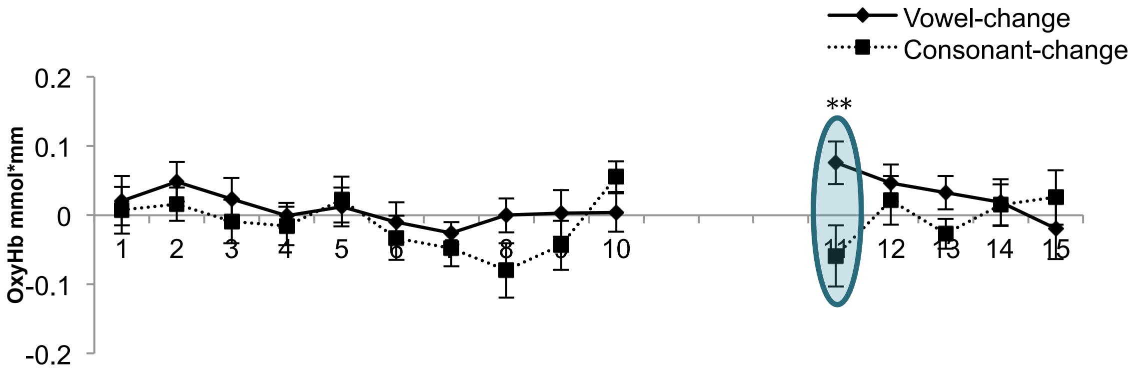

Figure 2.8 Experiment 2: Time courses of the relative hemodynamic changes (oxy-Hb) in the ***right frontal area.***
The dashed line depicts the results for the Consonant-change group; the continued line represents the Vowel-change group the group that heard a novel word in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed. There were significant differences between the two groups in the first block of the test phase **, p <0.001.

We also compared frontal, temporal, and parietal areas of the cortex in the left and right hemispheres using ANOVA[2]. In the first block of the test phase we observed a main effect of Group [F(1,31)=5.335; p < 0.05] due to greater overall brain activation for newborns in the Vowel-change group as compared to newborns in the Consonantal-change group. We found no main effect of Area [F(5,27)=0.306; n.s.] and no significant interaction between factors [F(5,27)=1.355; n.s.].

---

[2] As stated in the methods section, for each block, the channels showing artifacts were rejected. This is the reason why the degrees of freedom in the ANOVA vary from one comparison to another.

Moreover, in an ANOVA using the responses in the first block of the test minus responses in the last block of the exposure as the dependent measure, we found a main effect of Group [F(1,22)=7.839; $p < 0.01$] due to a general increment in the hemodynamic response from the familiarization to the test phase in the Vowel-change group, and a decrement in the hemodynamic response in the Consonant-change group (see Figure. 2.9). We found no main effect of Area [F(5,18)=1.566; n.s.], but a significant interaction between Area*Group factors [F(5,18)=2.870; $p < 0.05$]. Post hoc t-tests show a significant decrement in oxyHb concentration in the Consonant-change group and an increment in the Vowel-change group in the right frontal area (t(27)=4.387; $p < 0.0001$), and also in the left temporal (t(29)=2.324; $p < 0.05$) and the right parietal areas (t(28)=2.322; $p < 0.05$). The comparison in the last two areas, however, did not overcome the Bonferroni correction for multiple comparisons ($p < 0.0083$).

A binomial test showed that the majority of neonates displayed the above described activity in the right frontal area (increment in the hemodynamic response if included in the Vowel-change group, or decrement in the hemodynamic response if assigned to the Consonantal-change group), (binomial test, $p < 0.01$). The binomial test was also significant for the left temporal area (binomial test, $p < 0.01$), but not for the right parietal area (binomial test, $p = 0.05$).

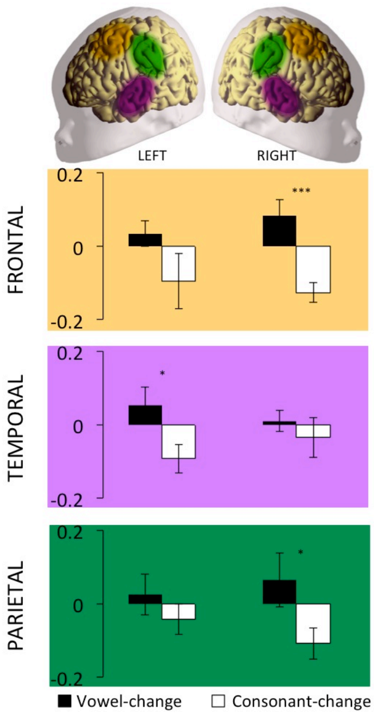

Figure 2.9 Experiment 2: oxy-Hb responses in the first block of the test phase minus responses in the last block of familiarization.
Channels located in the right frontal, left temporal and right parietal areas showed a decrease in oxy-Hb when participants heard a test word that shared the same vocalic-tier of the familiarization word but had different consonantal-tier (white bar). In contrast, there was an increment in the concentration of oxy-Hb from the familiarization to the test when neonates heard a word containing the same consonantal-tier of the familiarization word but novel vowels (black bar). Colored ellipses on the schematic neonate brain indicate the localization of the channels included in the areas of interest, ***, $p < 0.0001$; *, $p < 0.05$.

2.2.7 Discussion

In this study we found differential activations between the groups in the first block of the test phase. There was a significant decrease of oxy-hemoglobin when participants were tested with a word that contained novel consonants, and a significant increase in the hemodynamic response when the word contained novel vowels. We interpret these results in light of the brain activation observed in the previous experiment. This consisted in a decrease of oxy-hemoglobin when the familiarization word was presented again in the test phase, and an increase in the oxy-hemoglobin when a novel word (differing both in consonants and in vowels) was presented in the test.

Therefore, the activations observed in the current study rule out the first and second hypotheses proposed in the introduction. The results show that, despite their extraordinary phonetic discrimination capacities newborns seem to rely mainly on the most salient information available in the signal to recognize the familiar word. In particular, the data suggest that neonates, unlike adults and older infants, appear to rely on vowels or on some information carried by vowels to recognize a familiar stimulus.

These results have a number of implications for theories of speech processing, early memory, language acquisition and development. First and foremost the study evidences that newborn’s great discrimination capacities of phonemes (both for consonants and vowels), reported previously in various studies (e.g. Eimas et al., 1971) do not necessarily translate into a detail memorization of all the segments of a CVCV word. In the current study the consonants or vowels of the familiar and test words differed in as many phonetic features as it is possible in Italian (see section 2.2.2). Thus, it is very likely that our participants would have easily distinguished these words in a discrimination experiment. However, representing a word form relies both on the level of detail that newborns are able to perceive, and, critically, on their capacity to maintain a representation of this sound in memory. In fact our results suggest that despite newborns’ discrimination capacities they cannot memorize the information of both consonants and vowels for an interval of 2 minutes. The vowels of a word seem to be better encoded or at least they are easier to recognize after a delay, as compared to consonants[3].

At this point of our research it is not possible to clearly establish whether newborns also encode the consonants during the familiarization phase but are unable to recognize them in the test (possibly due to coarticulation) or whether they are disregarding consonantal information when learning the sound of a word, thus considering it just as noise. It seems, however, clear that the prevalence of vowels over consonants is not due to the incapacity to discriminate consonants.

It is premature to draw firm conclusions about whether the newborn's representations of words are structured in phonetic segments. The recognition of a familiar word at birth might be either based on the representation of the segmental qualitative properties of the vowels (height, i.e. vertical dimension; backness, i.e. horizontal dimension; roundedness, i.e. lip position), or on suprasegmental properties of vowels (pitch, length) or both[4]. However it is difficult to tell apart the quantitative and the qualitative aspects of vowels. In fact, front vowels such as /i/ are intrinsically pronounced with a higher pitch than back vowels such as /a/ (Taylor, 1933; Ewan, 1975). The changes in pitch are salient for newborns (Nazzi, Floccia, & Bertoncini,

---

[3] Although the results clearly showed that newborns were able to recognize the vowels of the word, the material does not allow us to evaluate whether newborns are sensitive to the vocalic-tier of the word or only to the content of vowels. In fact, in the current study we used simple words consisting of the same phoneme in the consonantal (s-s) and vocalic tiers (i-i). Thus, it is not clear whether newborns encoded only the identity of vowels (i.e., there were /i/ sounds), or also its sequence (i.e., */i/* follows */i/*). Familiarizing newborns with words composed of heterogeneous vocalic-tiers could be useful to address this question in future studies.

[4] In this experiment, the maximum changes in intrinsic phonemic pitch were indeed present in the Vowel-change condition, so newborns could have recognized the novel words in the test by means of the changes in this low-level information. However, the larger changes in the duration of phonemes were measured in the Consonant-change condition; thus if newborns represented only the quantitative information from words, this large acoustic change had to be detected by newborns especially considering that they are sensitive to consonant durational changes (Kushnerenko, et al., 2001). Participants did detect the novel word only in the Vowel-change condition, suggesting that the information available for the representation of words might be a combination of qualitative and quantitative aspects of the signal.

1998; Stefanics et al., 2009), just as vowel distinctions are (Cheour-Luhtanen et al., 1995; Aldridge, Stillman, & Bower, 2001; Cheour, et al., 2002), so both kinds of information could potentially support neonates' recognition of familiar sounds.

The observation that vowels are the main part of word-representations at birth is in agreement with previous behavioral studies suggesting that newborns pay more attention to vowels than to consonants. For instance, using a modification of the high amplitude sucking procedure, Bertoncini and colleagues showed that newborns familiarized with a set of syllables that share a phoneme (e.g. [*bi, ba, bo, be*] or [*bi, si, li, mi*]) cannot noticed the addition of a new syllable to the set when the novel syllable has a different consonant (e.g. [*da*] or [*di*]), whereas they notice the addition of a syllable when it has a different vowel (e.g. [*bu*] or [*ma*]) (Nazzi, Floccia, & Bertoncini, 1998; Stefanics et al., 2009), suggesting that early speech representations might not include in detail all phonetic segments, but mostly the vowels.

The neural effects showing significant changes in the hemodynamic responses were found in previously observed regions of the brain involved in language and memory processes, namely the left temporal, the right parietal, and the frontal areas; the former however, showed the most pronounced effects. The right frontal (Shallice et al., 1994; Rugg et al., 1996; Fletcher & Henson, 2001), and the parietal areas (Wagner et al., 2005; Cabeza et al., 2008; Olson & Berryhill, 2009) are usually activated when adult participants retrieve information from memory. Moreover, the right frontal area is activated in 3-month-old infants when they repeatedly listen to sentences but not when they listen to the same sentences played backwards (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002). Thus, the activation of these regions in newborns may be taken as an indication of the involvement of basic mechanisms of memory supporting the first stages of language acquisition.

The rightward response could also be related to the processing of supra-segmental information mostly carried by vowels. This interpretation conforms to a number of studies reported in the literature, reporting the right lateralization of prosodic and vowel processing in adults (Boemio, Fromm, Braun, & Poeppel, 2005; Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002; Meyer, Steinhauer, Alter, Friederici, & von Cramon, 2004) and infants (Homae, Watanabe, Nakano, Asakawa, & Taga, 2006; Telkemeyer et al., 2009). Empirical data and models of speech processing suggest that whereas fast spectral changes (i.e. segmental information) may enhance activation in the left hemisphere, prosodic contours elicit activation of the right hemisphere (Friederici & Alter, 2004; Zatorre & Belin, 2001). In sum, the right-lateralized response can be related to the processing of the information carried by vowels, and/or to specific mechanisms related to recognition capacities in newborns. Because prosodic processing and memory demands are both involved in our test, it is likely that the difference between the two conditions is triggered by a joint activation of these two processes.

These results therefore suggest that the consonantal bias for word learning and word processing should emerge during the first year of life. Newborns rely on the salient acoustic properties of vowels whereas twelve-month-olds start making use of consonants to establish the phonological representation of words (Hochmann et al., 2011). What could trigger this transition? While this intriguing developmental change requires further study, we explore some possible accounts.

First, the consonantal bias might emerge through brain maturation. Given that the right and left brain hemispheres develop at different rates (Thatcher, Walker, & Giudice, 1987), and that the cerebral blood flow has a right-hemispheric predominance in the first stages of development (Chiron et al., 1997), it is conceivable that some processes that mainly involve the right areas of the brain become functional earlier in development as compared to the processes involving the left-hemisphere. Consequently, when encoding a word-form, newborns may have access mainly to suprasegmental features, which are processed by the right hemisphere. Later in life, in part due to the greater maturation of the left-hemisphere, the more detailed segmental information might characterize better the long-lasting representations of words. This may also suggest that newborns are not yet extracting words from the speech input, but rather focusing on extracting structural properties cued by prosodic information (Morgan & Demuth, 1996a; Christophe, Mehler, & Sebastián-Gallés, 2001; Nespor & Vogel, 2007). For instance, from the variation of acoustic dimensions -including the percentage of vowels-, they may infer the mean length of words in their language; and from the prosodic contour they may also infer other syntactic properties such as word order (Nespor et al., 1996; Christophe et al., 2003).

Second, the consonantal bias might appear after infants have acquired the consonantal categories of their native language, that is, when they have shifted from phonetic to phonological representations (see Werker & Tees, 1984). Newborns are sensitive to a great number of phonemic contrasts (Eimas et al., 1971), by six months of age, however, infants' perception adjust to the vowel prototypes of the surrounding language (Kuhl, 2004; Kuhl et al., 1992; Polka & Werker, 1994) and by the end of the first year of life, infants also acquire native consonantal categories, while losing the sensitivity to consonantal contrasts absent from the language of exposure (Werker & Tees, 1984). Importantly, consonants are perceived more categorically than vowels (Pisoni, 1973), in consequence, after consonantal categories are defined words could be more reliably identified on the basis of their consonants because they provide a more stable representation of the sound of a word (Hochmann et al., 2011; Nespor et al., 2003).

Finally, the consonantal bias might also emerge as a result of statistical computations. By computing the relative frequency of consonants and vowels in the lexicon, the learner could realize that, being more numerous in most languages, consonants are more informative, and therefore rely more on consonants when learning new words (Keidel, Jenison, Kluender, & Seidenberg, 2007; but also see Bonatti, Peña, Nespor, & Mehler, 2007).

2.3 Conclusion of the Chapter

Previous studies on language acquisition uncovered precocious discrimination capacities in newborns. Our studies show that the toolbox with which humans are endowed to start learning about language also includes some basic representational mechanisms of speech sounds. In addition, we have shown that the newborn's discrimination capacities, reported in previous studies, do not necessarily translate into a detailed memorization of all the segments of a CVCV word. Newborns seem to represent mainly the elements that carry most of the energy within words, namely the vowels. These data contrast with previous results in older infants and adults who represent mostly the consonants of a lexical item. Given the differences in the way speech sounds are represented, it is likely that the learning mechanisms triggered early in life differ from the mechanisms use to process language later in development.

Chapter 3

Forgetting in newborns

By reviewing the vast literature on cognitive development and language acquisition -and perhaps only by observing infants in their course of development- one could easily conclude that infants are magnificent learners. It is during the first months and years that children learn and process massive amounts of new information. They learn the sounds of their mother tongue, the meanings of words and how to use those words in conversations. Because infants learn linguistic stimuli, one might have no doubts that they have the ability to store sounds and maintain the representation of these sounds in memory; we know that individuals would not be able to learn if their memory systems were not in place.

However, the motivation for carrying out the studies included in this chapter comes from the fact that, despite their basic capacities to remember information from words in the laboratory (see previous chapter), there is no evidence in the literature that before four months of age infants remember the most frequent words they heard in the surroundings (Mandel, Jusczyk, & Pisoni, 1995). One possible reason for these seemingly inconsistent facts might be that newborns' early experiences happen in surroundings that are very different from the silent environment used in laboratory studies. The continuous exposure to many sounds might prevent newborns from recognizing a previously heard sound in an everyday context, therefore limiting their basic abilities to remember words in the first months of life.

In this chapter, we explore the mechanisms that underlie forgetting in newborns, and the conditions under which, despite potential interfering factors, young learners can retain the sound of words in memory. We design the experiments to approximate better the normal acoustic environment of newborns; unlike our previous experiments in which we tested the participants in silent contexts, in the current studies we will incorporate noise at different phases of the experimental session. This manipulation should contribute to understand whether (and if so, how) newborns' memory for words withstand the presence of other sounds.

The reason to hypothesize that other stimuli present in the environment may disrupt the memory traces in newborns comes from the traditional theories of forgetting in adults, which we will briefly review in the following section.

Features of learning and forgetting in adults

As mentioned in Chapter 1, the first studies on forgetting in adults are attributed to the German psychologist Hermann Ebbinghaus. To study forgetting, Ebbinghaus compared the amount of time required to learn a list of nonsense words from scratch to the amount of time required to learn the same list the second time. He studied many different lists, separating learning and relearning sessions using retention intervals ranging from minutes to days. His findings showed that the *forgetting curve* generally fits a logarithmic function, with most forgetting happening soon after the information is learned and gradually slowing down over time. Although his methods were primitive, his basic premises have been confirmed and reaffirmed across a wide range of learning conditions.

Still, the fundamental question for many years has been what causes this pattern of forgetting. Traditionally, forgetting has been attributed to mechanisms such as trace decay or interference. The theory of decay poses that forgetting is due to the weakening of memories resulting from the passage of time. Some investigators favor the view that trace decay partially determines the loss of information for instance in short term memory (e.g. Baddeley et al., 1975). However, demonstrating that decay causes forgetting in adults is difficult, because it requires a proof that forgetting grows over time in the absence of other activities such as the storage of new experiences, thoughts or rehearsal; a condition that is obviously very difficult to achieve.

Another influential theory of forgetting put forward that memory can be disrupted because of the experience of other related events or percepts (Crowder, 1976; Anderson & Neely, 1996; Anderson, 2003). Because the number of experiences –and therefore the number of traces similar to the original one- increases over time, interference also provides a straightforward account of the forgetting curve. A number of qualitative distinct situations produce interference. For instance the storage of new experiences can interfere with remembering older ones (retroactive interference), but older memories can also impede learning of newer ones (proactive interference).

Forgetting is perhaps inevitable for many memories; however, in certain conditions it is be possible to minimize it. To begin with, the way a specific stimulus is encoded influences how much this stimulus will be successfully remembered later. For instance, the stimulus can be learned in massed presentations or in distributed presentations. Massed learning refers to a study time in which the stimulus to be remembered is not subject to any interruptions (e.g. intervening items). In contrast, learning is spaced or distributed when an intervening item or a long interval separates different study episodes of the same stimulus. A great number of studies have shown that memory can benefit from a distributed as opposed to a massive presentation of the information (for a review see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006).

Another important feature of memory that influences the rate of forgetting is its different vulnerability over time. New traces are initially more vulnerable to disruption than traces that have been kept for a longer time, because memories grow stable with time. The process of stabilizing memory traces is called consolidation (e.g. McGaugh, 2000). At least two types of consolidation have been proposed in the literature. Systems consolidation is used to refer to the process in which memories that are initially dependent upon the hippocampus undergo reorganization and become hippocampal-independent (see Alvarez & Squire, 1994 for a review). Estimates of the duration of systems consolidation vary, with some evidence suggesting that it might take hours to years in humans. Additionally, synaptic consolidation refers to the structural changes in the connections between neurons. These modifications rely on biological processes that may take minutes to hours to complete (see Dudai, 2004 for a review). The important issue here is that before those structural changes occur, the memory trace is vulnerable. Instead, after consolidation, a lower rate of forgetting is expected.

In sum, adults’ ability to remember a previously experienced stimulus can be influenced by competition arising from memories of similar events or stimuli. Moreover, forgetting in adults can be reduced by promoting a successful encoding of the information and by allowing the encoded material to be consolidated.

Whereas a great number of studies have focused on investigating the causes of forgetting in adults, the mechanisms that lead newborns to forget speech information are not yet established. Is it possible to show that in newborns (who most probably have less experiences accumulated than adults) the memory traces grow smaller only with the passage of time? Do newborns suffer from interference? If so, what kind of auditory experience produces interference over the sound of a word? Are the brain and its neural connections ready to consolidate word memories at birth? Can newborns benefit from the same learning strategies that enhance memory in adults?

It is possible that the mechanisms inducing forgetting verbal material are continuous throughout the lifespan, in which case newborns may show computations and limitations of verbal memory similar to those observed in adults. On the other hand, forgetting in adults may receive great influence from the development of top-down processes, for instance the development of semantic associations that cause interference, which, as stated in Chapter 1, are unlikely to affect newborns' memory.

In this chapter we will investigate some properties of forgetting in newborns. We will begin with an exploration of the interfering effects of auditory stimuli over the traces of a word (Experiments 3 and 4). In Experiment 5 we will examine the effects of presenting interfering sounds after silent periods of consolidation. Finally in Experiment 6 we will investigate the effects of distributed learning by introducing intervening sounds during the encoding phase. Altogether, the studies are designed to understand the contexts in which the presence of noise produce forgetting in newborns, and also when forgetting can be prevented or reduced. The ultimate question is whether the basic mechanisms of memory constrain the path of language acquisition, and if so how they shape language acquisition at birth.

3.1 Experiment 3: Recognition after intervening music

In Chapter 2 we have shown that newborns' brains have the capacity to retain the sound of words -or parts of it- during a silent 2-minutes retention interval. However, newborns' early experiences take place in surroundings that are very different from the silent environment used in our studies. Thus, in this experiment we investigate whether newborns' memory for words could withstand intervening auditory stimuli. We present an excerpt of instrumental music between familiarization and test. This task is clearly more demanding than the previous one, because in order to show a recognition response neonates not only need to maintain the trace throughout the retention interval, but they also have to overcome the possible interference caused by the intervening sound. A recognition response similar to the one observed in the previous studies would demonstrate that neonates were able to remember a word, and overcame the interference from music.

3.1.1 Participants

Twenty-eight healthy full-term neonates (15 females, mean age 2.8 days, range 1-5 days) participated in the experiment. Five neonates were tested but excluded from the analysis because head movements produced large motion artifacts (n=3), or because they cried before the end of the experiment (n=2). As in the previous experiments neonates were recruited at the newborn nursery at the Azienda Ospedaliera Universitaria Santa Maria della Misericordia in Udine, Italy. Participants were considered eligible if they had gestational ages between 38 and 42 weeks, Apgar scores ≥8 in the first and fifth minutes, diameter of head ≥33.5cm, no cephalohematoma, no think hair and normal history during pregnancy and delivery. Bioethics Committee of SISSA/ISAS (International School for Advanced Studies) approved the study; all parents gave informed consent prior to the experiment.

3.1.2 Stimuli

For this experiment we presented newborns with the same words used in Experiment 1 (*mita, pelu*). Moreover, an adapted excerpt of a Brahms' waltz, played on a piano, was used during the retention interval. Its duration was 10s (the same length of a block composed of 6 words) and had a mean intensity of 70dB.

3.1.3 Procedure

The experiment consisted of a familiarization phase, an interval, and a test phase (See figure 3.1). As in the previous studies the familiarization phase lasted 6 minutes and was organized in 10 blocks. Each block contained 6 identical words. Within blocks, words were separated by pauses of randomized length (0.5s or 1.5s), yielding blocks of approximately 10s each. Blocks were separated by time intervals of varying duration (25s or 35s) to avoid synchronization between stimuli occurrences and spontaneous oscillations. The test lasted for 3 minutes and it consisted of five blocks. In the test, half of the infants heard the same word that they heard in the familiarization, whereas the other half was presented with a novel token. A 2-minute interval separated the familiarization and the test phase, however, differently from previous studies, during the interval three blocks of an intervening instrumental music were inserted. The words used in the familiarization were counterbalanced across participants.

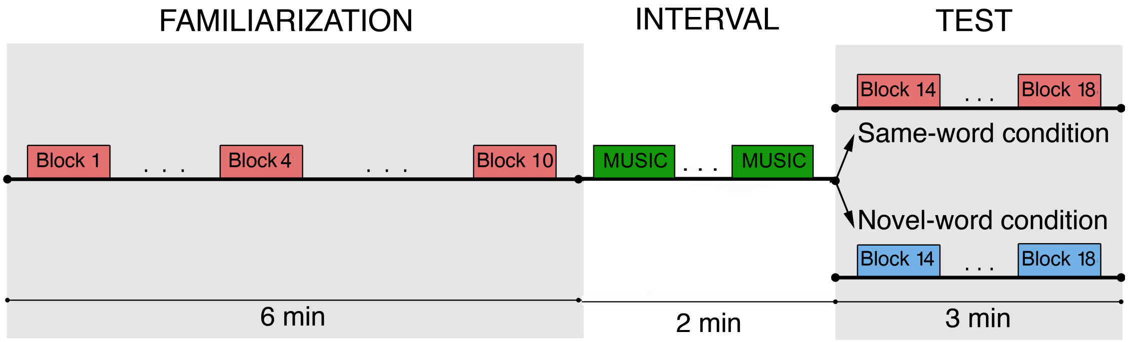

Figure 3.1Schematic diagram of the procedure used in Experiment 3.
During the familiarization phase, all the neonates were presented with 10 blocks composed of 6 identical words. During the interval, newborns heard three blocks of instrumental music. In the test, 5 blocks of the familiarization word were presented to half of the neonates whereas the other half listened to a completely novel word.

3.1.4 Data acquisition

The placement of the probes and NIRS data acquisition was identical to the one used in Experiments 1 and 2.

3.1.5 Data processing and analysis

The analyses performed were identical to those conducted for Experiments 1 and 2. In addition, to evaluate the recognition response, we conducted planned comparisons in the right-frontal area. This area was chosen based on the results of Experiment 2 and the evidence in the literature suggesting the involvement of the right frontal area in retrieval and recognition of information both in adults (Fletcher & Henson, 2001; Rugg et al., 1996; Shallice et al., 1994) and older infants (Dehaene-Lambertz et al., 2002)

3.1.6 Results

As in Experiment 1, significant differences in hemodynamic activity between conditions were found exclusively in the first block of the test phase (See figure 3.2); participants in the Novel-word condition showed greater concentrations of oxyHb than participants of the Same-word condition (permutation test, $p<0.01$). Using ANOVA we also compared the oxyHb concentration changes in the first block of the test, in frontal, temporal, and parietal areas of the cortex in the left and right hemispheres. We found a main effect of Group (ANOVA, [F(1,26) = 19.318; $p<0.0001$]), due to a general greater response in the Novel-word Group as compared to Same-word group. There was no main effect of Area [F(5,17)=0.201; n.s.] and no significant interactions between factors [F(5,17)=0.463; n.s.]. Moreover, as in the previous study, participants assigned to the Same-word and Novel-word conditions did not differ during the familiarization phase.

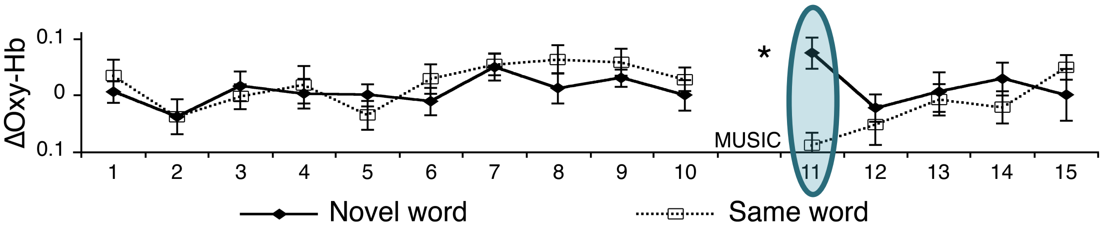

Figure 3.2 Time courses of the relative hemodynamic changes (oxy-Hb) ***averaged across all channels*** in Experiment 3.
The dashed line indicates the time series for the group that heard the same word before and after the pause; the continued line represents the group that heard a novel word in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed. There was a significant difference between the Same-word and Novel-word groups in the first block of the test phase * p<0.01

Furthermore, changes in oxyHb concentration from the last block of the familiarization to the first block of the test differed between the two conditions (permutation test, *p*<0.01; t-test t(23)= -3.253, *p*<0.005). Participants in the Same-word condition showed a decrease in concentration of oxyHb between familiarization and test, while participants in the Novel-word condition showed increased hemodynamic responses.

We also performed planned comparisons in the first block of the test phase in the right frontal area. The permutation test showed significant differences between the Same-word and Novel-word conditions (permutation test, *p*<0.001) (See figure 3.3). Moreover, there were significant variations in oxy-Hb concentration between the last familiarization block and the first block of the test phase (permutation test, *p*< 0.05; t-test, t(23)= -2.904, *p*<0.01). Additionally, there were significant differences between the groups in the first block of the test phase in the left frontal (permutation test, *p*< 0.05), right temporal (permutation test, *p*< 0.05) and right parietal areas (permutation test, *p*< 0.01), however these comparisons did not overcome the corrections for multiple comparisons.

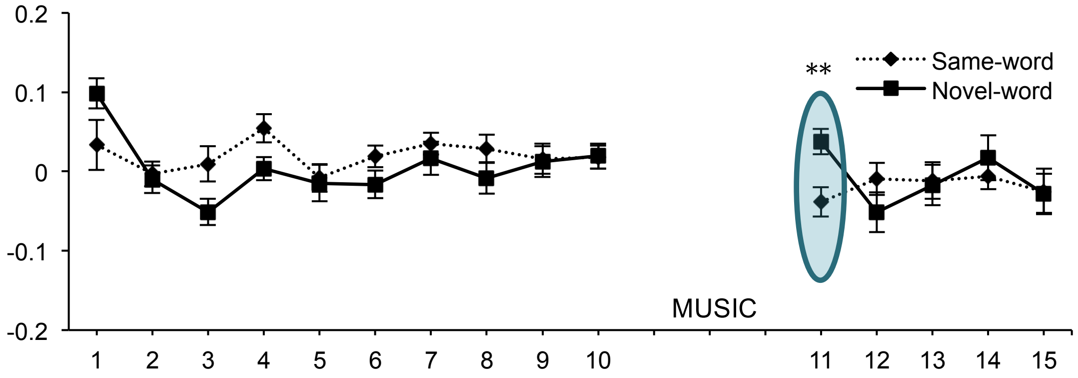

Figure 3.3 Experiment 3: Time courses of the relative hemodynamic changes measured in ***the right frontal area.***
Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed. There was a significant difference between the groups in the first block of the test phase ** p<0.001

3.1.7 Discussion

The differences observed between the two groups in the first block of the test phase confirm the findings of the previous experiments, namely that newborns are able to distinguish the familiar and the novel word after a retention interval of a few minutes. In addition, this study shows that newborns remembered the familiar word despite the presence of noise during the retention interval. This implies that newborns can retain the sound of a word for some minutes in at least two contexts: 1) silent environments (see chapter 1), and 2) in the presence of sounds that are frequently present in the newborn's environment i.e. lullabies.

This result, however, cannot be taken as evidence that the traces of a word are generally resistant to the presence of other sounds in the surroundings. It is possible that newborns retain the sound of a word only in some, but not in all noisy contexts. For instance, newborns' representation of a word may not withstand stimuli that are more similar to the stimulus that originated the trace. In fact, in adults it is frequently observed that not every type of intervening experience impairs memory, or at least not to the same extent. The experience needs to be similar to the previously learned one, causing competition between the two memories (e.g. Barnes & Underwood, 1959; Deutsch, 1970; Morton & Chambers, 1976). For example, in laboratory tasks, if participants are required to recall the word "orange" when presented with the cue word (i.e. fruit), they accidentally retrieve the word banana after having practiced the pair fruit-banana in the laboratory (Anderson, Bjork, & Bjork, 1994). In the same way, when searching for today's parking space, people usually fail to retrieve the correct location of the car and confuse it with the place where they parked yesterday or the days before, because the memory of today's parking place gets interference from other similar memories.

However, one may ask: what count as similar stimuli for newborns? If newborns' word representation does not suffer interference from a music excerpt, it might be because newborns store words in a format that does not overlap with the representation of music stimuli. In fact, early in life and until adulthood, the human brain processes speech and music in different ways, recruiting distinct neural networks (Zatorre, Belin, & Penhune, 2002; Peretz & Coltheart, 2003; G. Dehaene-Lambertz et al., 2010; Perani et al., 2010; Kotilahti et al., 2010). The processing of speech and music as two distinct auditory experiences might have allowed newborns to retain words in the presence of music.

It is, however, possible that newborn's memory for single sounds is resistant to any intervening sound. Unlike adults, young infants have stored few memories; these memories may not interfere to one another until the limits of the storage system would be reached. It is also possible that the presence of sounds that are more similar to the trace would lead newborns to forget the familiar word.

In Experiment 4 we will further explore the newborn's ability to maintain the representation of word sounds in noisy environments. In particular we will ask whether newborns' representation of words endure the presence of other speech sounds in the surroundings.

3.2 Experiment 4: Recognition after intervening words

In this experiment we evaluate newborns' memory for a word when another word sound intervenes before the test. We hypothesize that forgetting might be higher when the initial encoding event is followed by similar auditory experiences, in particular that speech stimuli might cause greater interference over the trace of a word than the music used in Experiment 3. Alternatively, newborns' representation of words might endure the presence of any sound in the environment including speech. Such result would suggest that newborns' representations of the sound of words are durable. On the other hand, differences in the amount of interference caused by an intervening word and by music can give insights into how stimuli in these two auditory domains are represented in the newborn brain.

3.2.1 Participants

Twenty-eight healthy full-term neonates (12 males, mean age 2.9 days, range 1-5 days) were included in the analysis. Eight additional neonates were tested but excluded from the analysis because of fussiness or because crying before the end of the experiment. Neonates were recruited at the newborn nursery at the Azienda Ospedaliera Universitaria Santa Maria della Misericordia in Udine, and the eligibility criteria used was the same as in previous experiments. Bioethics Committee of SISSA/ISAS (International School for Advanced Studies) approved the study; all parents gave informed consent prior to the experiment.

3.2.2 Stimuli

In this experiment we used three pseudowords (*mita, pelu, noke*). A female, native speaker of Italian recorded the words. All the words were pronounced using child directed speech, carried first syllable stress, had a CVCV (consonant-vowel-consonant-vowel) structure, and were edited to have the same intensity (70dB) and duration (700ms). The pseudoword *noke* had mean pitch of 258Hz (range 206-438Hz), *mita* and *pelu* were the same words used in previous studies (see section 2.1.2).

3.2.3 Procedure

The procedure was exactly the same as that used in Experiment 3, except that instead of inserting three blocks of instrumental music during the interval, we inserted three blocks of an intervening word (see figure 3.4).

3.2.4 Data acquisition

The placement of the probes and NIRS data acquisition was identical to the one used in the previous experiments.

3.2.5 Data processing and analysis

The analyses performed were identical to those conducted for Experiments 3.

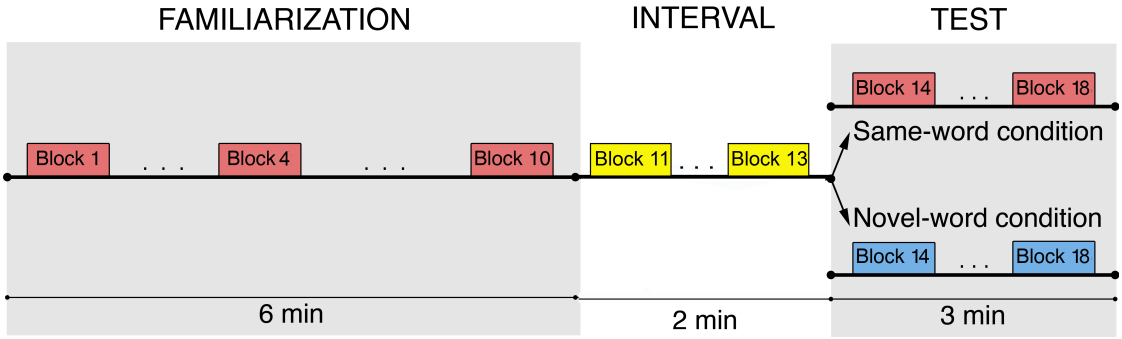

Figure 3.4 Schematic diagram of the procedure used in Experiment 4.
During the familiarization phase, all the neonates were presented with 10 blocks composed of 6 identical words. During the interval, newborns heard three blocks of an intervening word. In the test, 5 blocks of the familiarization word were presented to half of the neonates whereas the other half listened to a completely novel word.

3.2.6 Results

As in the previous studies, we found no significant differences during the familiarization phase between the newborns familiarized with the words *pelu* or *mita* (permutation test, $p = 0.18$). Moreover, in the responses averaged across all the channels and subjects per Condition we found no significant differences between the Same-word and Novel-word conditions in the test phase (permutation test, ps > 0.21; see Figure 3.5).

For the channels located in the right-frontal area, we found no significant differences between the conditions in the familiarization phase (permutation tests; p = 0.14). In addition, the permutation test in the first block of the test phase showed no significant difference between the Same-word and Novel-word conditions (permutation test, p=0.30) (see figure 3.6).

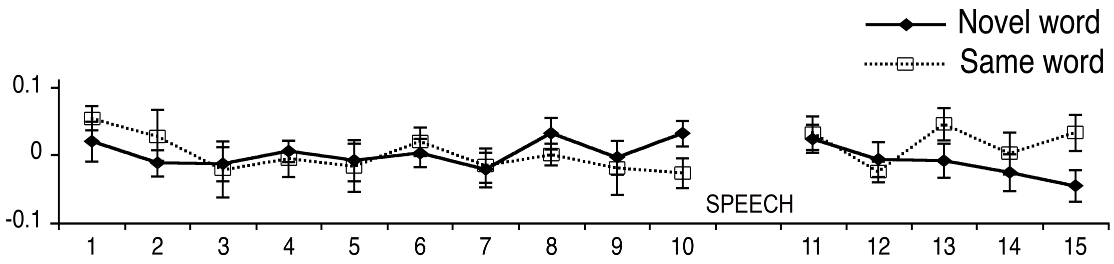

Figure 3.5 Time course of the relative hemodynamic changes in Experiment 4
Data were averaged across all the channels and subjects per group. There were no significant differences between the groups. The dashed line indicates the time series for the group that heard the same word before and after the pause; the continued line represents the group that heard a novel word in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed.

We also performed an ANOVA to compared brain hemodynamic responses in six regions: frontal, temporal, and parietal regions of the left and right hemispheres. We did not observe any main effect (Condition, F(1,24)=0.548; n.s., Area, F(5,21)=1.193; n.s.], and no significant interactions between factors [Condition*Area, F(5,21)=0.810; n.s.]. We also failed to observe significant variations in oxy-Hb concentration between the last familiarization block and the first block of the test phase (permutation test, $p > 0.50$).

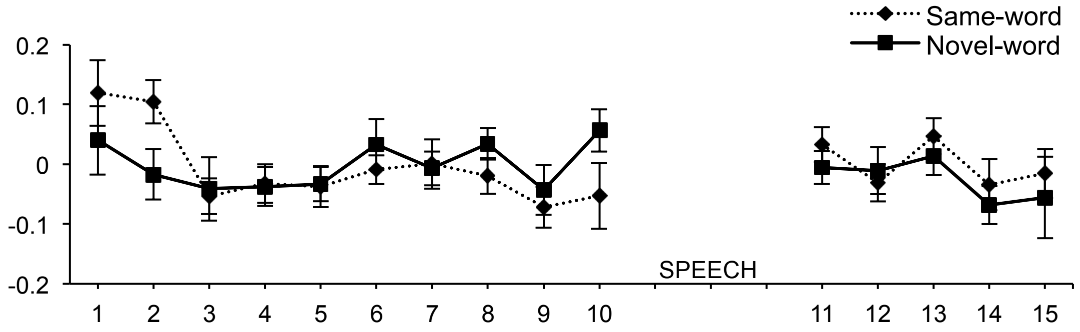

Figure 3.6 Experiment 4: Time courses of the relative hemodynamic changes ***in the right frontal area.***
Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed. We found no differences between the two groups.

3.2.7 Discussion

In contrast with the results obtained in the previous experiment, in this experiment we found no evidence of recognition. These results cannot be accounted for by theories of decay, since the time elapsed between the familiarization and the test phase was exactly the same as in the previous studies of this thesis, were recognition was verified. Instead, it is likely that the presence of another word during the retention interval disrupted newborns' ability to retain the previously learned word. It seems that, similar to adults, newborns' memory is fragile when it is followed by other related experiences (see Crowder, 1976; Anderson, 2003 for reviews on interference in adults).

Our results suggest that speech sequences in newborns are encoded in a format that does not generalize to music sequences. An interesting area for speculation constitutes the level over which the newborn brain computes similarity between auditory stimuli. As a first alternative, it is possible that similarity is computed with reference to low-level acoustic cues. The word presented during the familiarization and retention interval had the same pitch, duration, intensity, voice quality, syllabic structure, and stress pattern. In contrast, the instrumental music had a more complex melodic contour and continuous transitions. Thus it is possible that the degree of acoustic similarity between the familiarization word and the intervening word determined the amount of interference.

A second alternative is that a conspecific bias would be responsible for the contrastive effects of music and words over the memory trace. At birth infants display preferences towards biological stimuli, and in particular for conspecifics (Bonatti, Frot, Zangl, & Mehler, 2002); for instance, neonates prefer to look at point-like displays that follow biological movement over non-biological motion (Simion, Regolin, & Bulf, 2008), they show a visual preference for human faces (Johnson, Dziurawiec, Ellis, & Morton, 1991; Valenza, Simion, Cassia, & Umiltà, 1996; Cassia, Simion, & Umiltà, 2001) and privilege natural speech stimuli over other environmental sounds (Vouloumanos & Werker, 2004). The interference speech-speech could have thus arisen from a mechanism to detect conspecifics, which is not triggered by the music stimuli.

A third alternative is that interference arises from the similarity of the sources that produced the auditory stimuli. In our experiments, words (both the familiar and the intervening) were generated by the vocal tract, whereas the melody was produced by an instrument. Thus, similarity in the neonate brain could have been computed with reference to this property of the stimuli. This alternative is supported by studies showing that vocal and non-vocal sounds are processed by different areas of the adult brain (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Belin, Zatorre, & Ahad, 2002; Belin, Fecteau, & Bédard, 2004).

Finally, similarity between music and speech might be computed at abstract levels, based on prewired mechanisms that are specialized to process these two types of stimuli separately. This possibility would be supported by theories that propose a modular-like system of the human mind, which has evolved to process and encode speech stimuli and music separately (Fodor, 1983; Peretz & Coltheart, 2003). Additional support for this alternative comes from studies showing that the infant brain is specialized to process speech and to process music since birth (Peña et al., 2003; Perani et al., 2010, 2011; Trehub, 2001, 2003), and that these two sounds continue to be processed differently in the infant (Dehaene-Lambertz et al., 2010) and the adult brain (Zatorre et al., 2002).

Another issue raised by the results of the current study concerns the features of the information storage system in newborns. Based on interference theories and theories of consolidation in adults, one could hypothesize that the intervening word produces interference because it is presented immediately after the familiarization phase, when the trace is still unstable. Perhaps, the newborn’s brain requires long silent periods -or at least periods of no intervening speech stimulation- for the stabilization of the trace. This alternative is sustained by previous research showing that the process of stabilization (called consolidation) takes some time after learning because it involves structural changes in the synaptic connections between neurons (Dudai, 2004). This would suggest that forgetting might not be induced always after hearing an intervening word, but only when newborns are exposed to this intervening word immediately after the familiarization. This account would predict that after consolidation, the presence of an intervening word would not cause interference, or at least not to the same degree. The next experiment was designed to explore this alternative.

3.3 Experiment 5: Time for consolidation

In the previous experiment we found that newborns forget the familiarization word as soon as they hear other speech stimuli. In this study we investigate whether this phenomenon can be overcame by allowing the encoded material to be consolidated. For this purpose the experimental paradigm was modified so that the intervening word was not presented immediately after the familiarization phase, but rather 2 minutes later. If participants fail in this experiment, then this would suggest that hearing a second word is sufficient to disrupt newborns' recognition of the familiarization word, independently of the moment in which it is presented. If on the other hand the familiarization word was recognized in the test, this would indicate that the absence of other speech stimuli, minutes after the familiarization, favors the consolidation of the traces.

3.3.1 Participants

Forty newborns (13 males; mean age: 2.8 days, range: 2-5 days) participated in this experiment. Six more babies were tested, but were not included in the analyses, because they failed to complete the experiment due to crying, we also excluded infants because of experimental error (n=2), because of parental interference (n=1) or because of the amount of data rejected in the artifact rejection phase was greater than threshold (n=1). Neonates were recruited at the newborn nursery at the Azienda Ospedaliera Universitaria Santa Maria della Misericordia in Udine, Italy. Eligibility criteria were the same as in previous experiments. Parents gave informed consent prior to the experiment. The Ethics Committee of SISSA, granted permission.

3.3.2 Stimuli

Stimuli were the same as those used in Experiment 4.

3.3.3 Procedure

The procedure was similar to the one used in Experiment 4, except that the three blocks of the intervening word were inserted two minutes after the end of the familiarization phase (rather than immediately after). As a consequence, the retention interval that separated the familiarization and the test phase lasted 4 minutes (see figure 3.7).

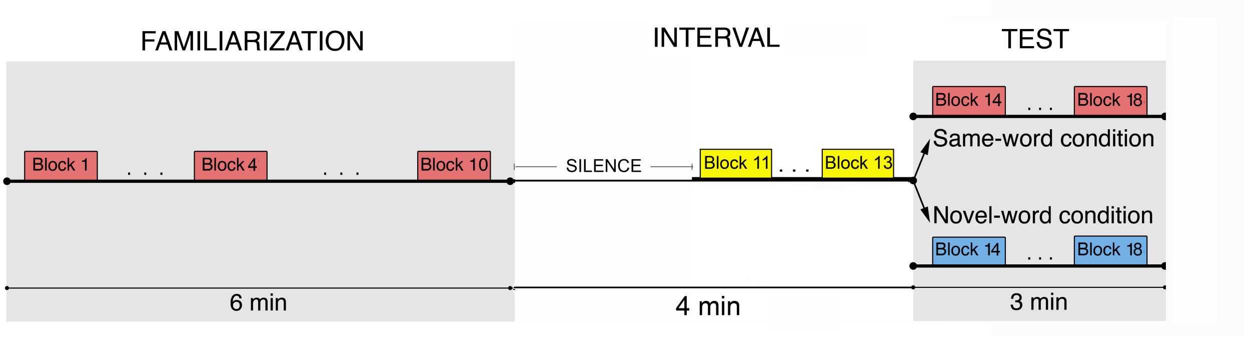

Figure 3.7 Schematic diagram of the procedure used in the Experiment 5.
During the familiarization phase, all the neonates were presented with 10 blocks composed of 6 identical words. During the interval, newborns heard silence for 2 minutes, and then three blocks of an intervening word were presented. In the test, 5 blocks of the familiarization word were presented to half of the neonates whereas the other half listened to a completely novel word.

3.3.4 Data acquisition

The placement of the probes and data acquisition was identical to those used in previous experiments.

3.3.5 Data processing and analysis

The analyses performed were identical to those conducted for Experiment 4

3.3.6 Results

As in the previous studies, we found no significant differences during the familiarization phase (permutation test, $p = 0.20$). Moreover, in the responses averaged across all the channels and subjects per condition we found no significant differences between the Same-word and Novel-word conditions in the test phase (permutation test, ps > 0.09; see Figure 2.8). In additional analysis including all the participants from both conditions, we subtracted the responses in block 11 minus the responses in block 10. The mean value of this calculation was positive (mean: 0.03 mmol*mm; standard error: 0.01) and was significantly different from baseline ($p<0.05$). This shows that there was a significant increment in the hemodynamic response from the familiarization to the interfering phase.

In an ANOVA, we compared brain hemodynamic responses in the first block of the test phase, in six regions: frontal, temporal, and parietal regions of the left and right hemispheres. We observed a main effect of Condition [ANOVA, $F(1,26)=8.690$; $p < 0.01$] due to the greater hemodynamic responses of neonates in the Novel-word condition as compared to neonates in the Same-word condition. There was no main effect of Area, [ANOVA, $F(5,22)=0.667$; n.s.], and no significant interactions between factors [ANOVA, $F(5,22)=1.516$; n.s.].

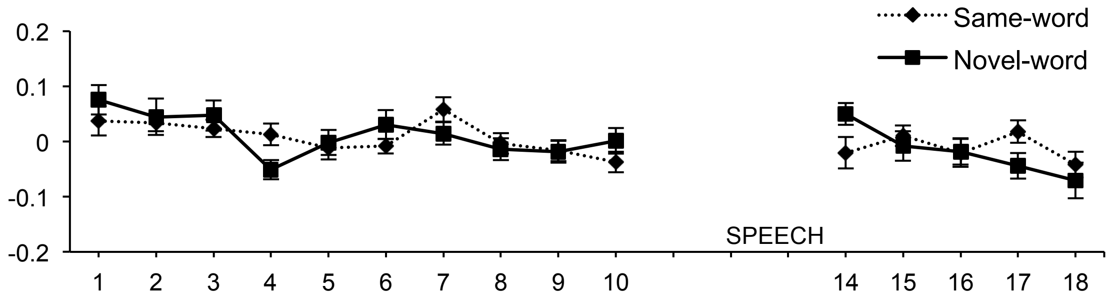

Figure 3.8 Time courses of the relative hemodynamic changes ***averaged across all channels*** in Experiment 5.
There were no significant differences between the conditions in any block. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed.

Additionally, based on the previous studies we performed planned comparisons in the first block of the test phase in the right frontal area (Figure 3.9). Significant differences in brain activation between the two conditions were observed in this area (permutation test, *p*<0.001). No other block showed significant differences in brain activity between neonates in the Novel-word and in the Same-word conditions.

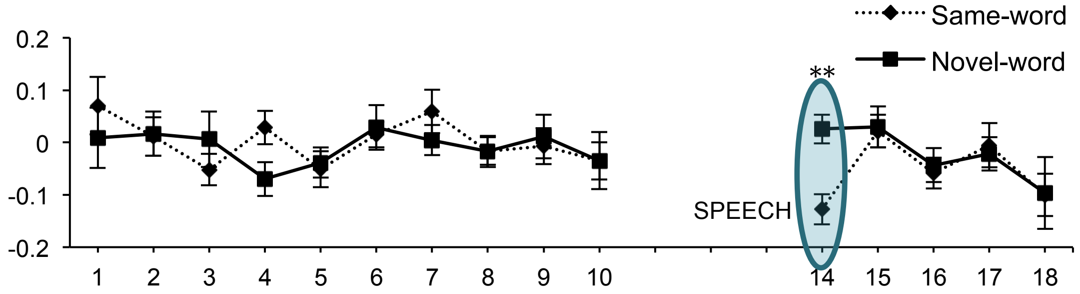

Figure 3.9 Experiment 5: Time courses of the relative hemodynamic changes ***in the right frontal area.***
Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed. We found significant differences between the conditions in the first block of the test phase **, p <0.001.

In the right frontal area, the analysis of the responses of the first block of the test as compared to the last block of the familiarization showed a decrement of oxyHb in the Same-word condition and an increment in the Novel-word condition, the differences between the changes in concentration in the two conditions was marginally significant (t-test, t(23)=-2.142, $p<0.05$; permutation test $p= 0.09$).

3.3.7 Discussion

The results indicate that newborns' brains recognized the familiarization word and distinguished it from the novel word in the test. Therefore, when newborns listen to speech stimuli some minutes after the familiarization, these stimuli do not interfere with the memory of the familiarization word. This suggests that in the previous experiment newborns failed to show a recognition response not because a different word intervened during the retention interval, but because it intervened at the moment in which the familiarization word was not consolidated. Instead, in the current experiment the trace of the familiarization phase became more stable, possibly because of the synaptic consolidation that took place within the firsts minutes after the encoding phase. It thus seem that in the absence of other speech stimuli, the newborn brain is able to consolidate the trace of a recently learned word, and therefore transforms the sensory experience into a representation resistant to interference. These results closely resemble the consolidation effects often found in adult and animal studies (Muller and Pilzecker as cited for instance in Lechner, Squire, & Byrne, 1999; Dudai, 2004; Baddeley, Eysenck, & Anderson, 2009).

These results, thus suggest not only that certain mechanisms of consolidation are active throughout the lifespan, but also that they can be crucial to promote learning in the presence of interfering sounds.

In addition, this study revealed some properties of the human's speech storage system in its initial state. In particular, given that the trace lasted at least 4 minutes before decaying, it is clear that the representational system at birth is not limited to the traces left by the sensory experience. In fact, results from a previous study suggest that the time span of sensory auditory memory in neonates decays very rapidly (Cheour et al., 2002). In their study, Cheour and colleagues showed that newborns' mismatch negativity response or MMN (a very well-known event-related potential component associated to the experience of a sound that differs from a recently experienced and repeatedly heard stimulus) was elicited when the deviant sound was presented 800ms after the standards. The authors reported no MMN in newborns with longer stimulus-onset asynchronies, and suggested that the time span of sensory memory in newborns decays in less than 1500ms. Based on this study, and given that this time span is considerably shorter than the retention interval used in our study, it may be reasonable to assume that newborns are able to form a more stable representation than that stored in sensory memory. Moreover, as opposed to the sensory –also called echoic– memory that retains the most recently heard sound, the memory system that was evidenced in this experiment retained the familiarization word despite of the fact that infants heard other sounds before the test.

It is an open question whether newborns are able to maintain the representation of the familiarization word for longer retention intervals than the one tested in the current study. Systems consolidation –which supposedly mediates the retention of information in long-term memory– takes days to months to complete; it usually depends on the hippocampal formation and some other medio-temporal structures that undergo substantial postnatal developmental change throughout infancy and into the school years (e.g. Bauer, 2006; Richmond & Nelson, 2007). The slow maturation of the hippocampal areas may limit the consolidation of early memories. However, additional studies must be carried out in order to understand whether –independently of the hippocampus– the circuits that have originally encoded the information, could consolidate early speech memories, and whether the representation of word-sounds could be maintained in this system for longer periods. This possibility arises from previous studies that have evidenced that consolidation can take place outside the hippocampus, in the same neural circuits that have originally encoded the information (Brashers-Krug, Shadmehr, & Bizzi, 1996; Shadmehr & Holcomb, 1997).

Additionally, we might ask whether the phenomena described so far in this thesis is representative of the mechanisms that play a role in real language acquisition situations. In a normal learning context, the sound of a word is not repeated alone but rather presented together with other word-sounds in the speech. In order to identify the variables that modulate the learning process and because of methodological restrictions, our studies have been forced to test highly constrained paradigms and materials, substantially different from the newborn's common environments. Because our interest is also to understand newborn's memory capacities in ecologically plausible situations, in the next experiment we modify the testing paradigm attempting to approximate a newborn's everyday exposure.

3.4 Experiment 6: Effects of intervening words during encoding

In this experiment we test newborns' recognition capacities in interfering conditions. Crucially and differently from the previous experiments of this Chapter, the interfering stimuli will be presented during the encoding phase rather than during the retention interval. This simple manipulation allows us to systematically assess different questions. First, can the newborn brain retain the sound of words in contexts that resemble better their everyday exposure to speech? Generally, in the newborn's environment, words are not isolated but rather embedded in simple sentences containing other words. Would these intervening words interfere with the memory of the most frequently presented word? Second, by incorporating exactly the same type and amount of interfering sounds that were introduced during the retention interval in the previous experiments, we can ask whether the interference effects are modulated by the quantity and quality of the intervening words, or rather affected by the moment in which participants heard this sound. Third, this modification allows us to explore the effects of distributed learning upon retention of verbal information at birth, contributing to elucidate whether, similar to adults, newborns can benefit more from distributed presentations than from massed familiarization episodes.

3.4.1 Participants

Forty newborns (14 males; mean age: 2.7 days, range: 2-5 days) participated in this experiment. Five additional infants were tested, but were not included in the analyses, because they failed to complete the experiment due to fussiness. Neonates were recruited at the newborn nursery at the Azienda Ospedaliera Universitaria Santa Maria della Misericordia in Udine, Italy. Eligibility criteria were the same as in previous experiments. Parents gave informed consent prior to the experiment. The Ethics Committee of SISSA, granted permission.

3.4.2 Stimuli

Stimuli were the same used in Experiments 4 and 5.

3.4.3 Procedure

The experiment consisted of a familiarization phase, a silent interval, and a test phase (see figure 3.10). The familiarization phase lasted 6 minutes and was organized in 10 blocks. Each block contained 8 words: 6 identical words (either *mita* or *pelu*) and 2 intervening words (*noke*) presented in random order. Within blocks, words were separated by pauses of randomized length (0.5s or 1.5s), yielding blocks of approximately 13s each. Blocks were separated by time intervals of varying duration (25s or 30s) to avoid synchronization between stimuli occurrences and spontaneous oscillations. A 2-minute silent interval was inserted between the end of the familiarization and the beginning of the test phase. The test lasted for 3 minutes and it consisted of five blocks. In the test, half of the infants heard the most frequent word that they heard in the familiarization whereas the other half was presented with a completely novel token. Test blocks contained 6 identical words each.

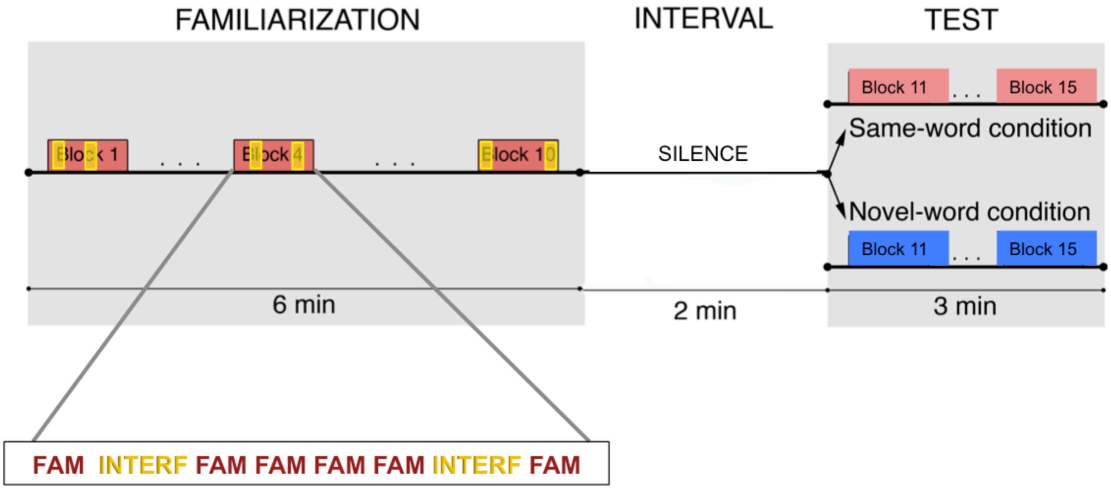

Figure 3.10 Schematic diagram of the procedure used in the Experiment 6.
During the first phase, all the neonates were familiarized with 10 blocks composed of 6 identical words (FAM) and 2 interfering words (INTERF) presented in random order within each block. In the test, 5 blocks of the familiarization word were presented to half of the neonates whereas the other half listened to a completely novel word. A 2-minutes silent retention interval separated the familiarization and the test phases.

3.4.4 Data acquisition

The placement of the probes and data acquisition was identical to the one used in previous experiments.

3.4.5 Data processing and analysis

The analyses performed were identical to those conducted for previous experiments.

3.4.6 Results

In the responses averaged across all the channels and subjects per Condition we found no significant differences between the Same-word and Novel-word conditions in the familiarization phase (permutation test, $p > 0.30$). In the test phase we found significant differences between the conditions in the first block (permutation test, $p<0.05$), and also in the second block of the test phase (permutation test, $p<0.05$).

We also performed planned comparisons in the first block of the test phase in the right frontal area, where we found significant differences between the two conditions (permutation test, $p<0.001$) (see Figure 3.11). No other block showed significant differences in brain activity between neonates in the Novel-word and in the Same-word conditions. Moreover, differences in the responses from the first block of the test and the last block of the familiarization showed a decrement of oxyHb in the Same-word condition and an increment in the Novel-word condition; differences between the changes in concentration in the two conditions was marginally significant (permutation test $p= 0.06$).

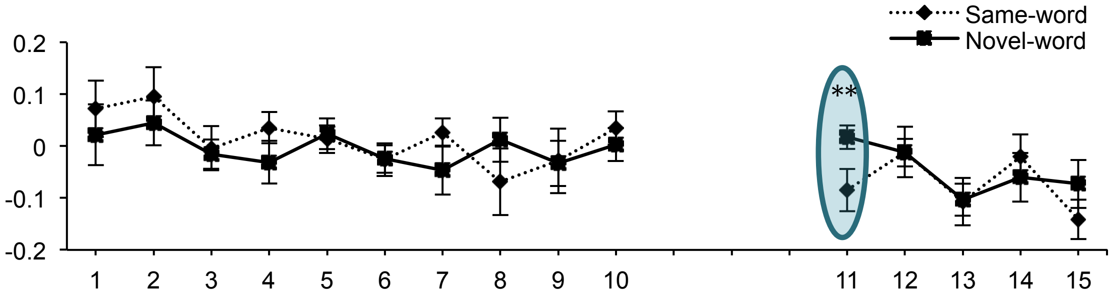

Figure 3.11 Experiment 6. Time courses of the relative hemodynamic changes measured in ***the right frontal area.***
The dashed line indicates the time series for the group that in the test heard the familiarization word; the continued line represents the group that heard a novel word in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed. There was a significant difference between conditions in the first block of the test phase. **, p <0.001.

Additionally, significant differences in brain activation between the two conditions were observed in the second block of the test in the left temporal (permutation test,

$p<0.01$), and right parietal areas (permutation test, $p<0.01$) (see Figures 3.12 and 3.13).[5]

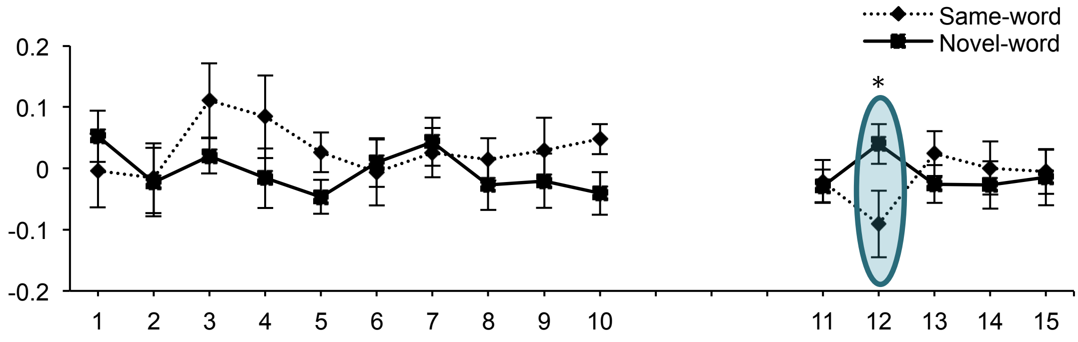

Figure 3.12 Experiment 6: Time courses of the relative hemodynamic changes measured in the ***left temporal area.***
The dashed line indicates the time series for the group that in the test heard the familiarization word; the continued line represents the group that heard a novel word in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed. There were significant differences between conditions in the second block of the test phase. *, p <0.01.

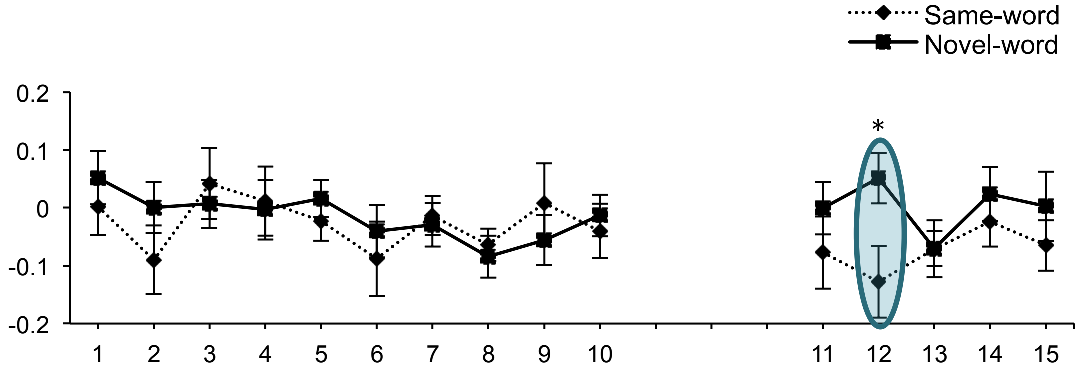

Figure 3.13 Experiment 6: Time courses of the relative hemodynamic changes measured in the ***right parietal area.***
The dashed line indicates the time series for the group that in the test heard the familiarization word; the continued line represents the group that heard a novel word in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed. There were significant differences between conditions in the second block of the test phase. *, p <0.01.

---

[5] Despite that none of these analyses overcome Bonferroni corrections for multiple comparisons, we report these differences because they can be informative in future studies.

Finally, a linear regression analysis computed over the time course of the experiment showed a significant reduction in the hemodynamic response in the left anterior areas (y=-0.01x + 0.05; coefficient 1 p = 0.002; coefficient 2 p = 0.003; R^2=0.671). Figure 3.14 illustrates the changes in oxyHb concentration over the familiarization and test blocks of the experiment. For statistical purposes, we compared the beginning (first five blocks) and the end (last five blocks) of the familiarization phase using t-test. This comparison further showed that the hemodynamic response changed throughout the experiment, t(4)=3.006, p<0.05.

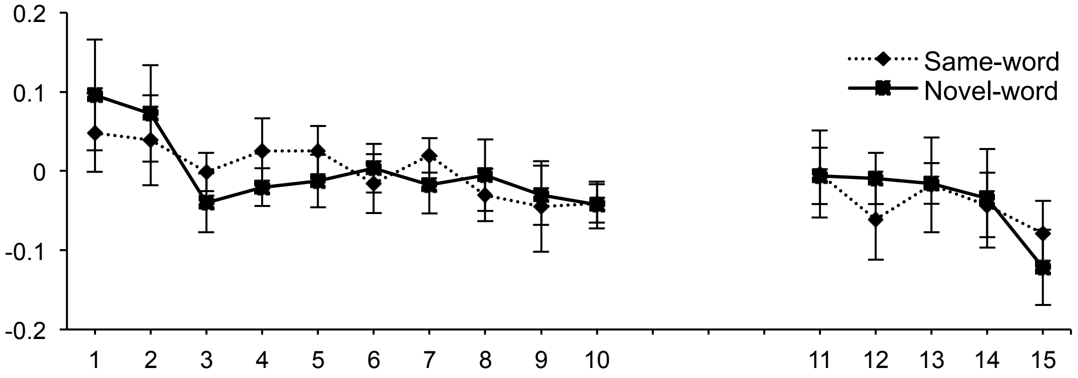

Figure 3.14 Experiment 6. Time courses of the relative hemodynamic changes measured in the ***left frontal area.***

There was a significant reduction in the response throughout the experiment. The dashed line indicates the time series for the group that in the test heard the familiarization word; the continued line represents the group that heard a novel word in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed.

3.4.7 Discussion

These results indicate that the neonate brain is able to remember a word in contexts that resemble everyday situations. More specifically, it is able to retain the information of a frequently heard word, even if it is presented in concomitance with other words during the familiarization.

This result together with the results of Experiment 4 shed light on the mechanisms of forgetting in newborns. In Experiment 4, the presence of an intervening word disrupted the newborn's ability to retain the familiarization word. In the current experiment, exactly the same intervening sound, presented the same number of times, did not cause interference. There are different reasons that may explain why the same amount and type of information caused forgetting in the former situation whereas in the later one favored recognition. First, in the current experiment, the intervening words were presented during the encoding phase, whereas in Experiment 4 the words were presented during the consolidation interval. In the current experiment, thus, the familiarization word presumably was consolidated (because it did not received interference in the first minutes immediately after familiarization) whereas in Experiment 4 the sound was forgotten because the interfering words impeded consolidation.

Another crucial factor that may explain why newborns succeeded in Experiment 6 concerns the structure of the familiarization phase. The results of the current experiment show that, similar to adults, newborns may benefit from distributed familiarization episodes. The recognition response observed here suggests that the reactivation of memories during the encoding phase (obtained from the interspersed presentation of familiarization words and interfering sounds) can protect newborns' memories against forgetting. This result strongly parallels traditional studies (see Ebbinghaus, 1964) and recent evidence showing the advantages of distributed learning as compared to massed learning in adults (Madigan, 1969; Crowder, 1976; Cepeda et al., 2006).

Although it may seem counterintuitive, the variability of the input during encoding could have also contributed to create a more robust representation of the familiarization word in the present study. The presence of more than one sound may benefit word recognition because it helps infants identifying contrastive aspects of the speech items, consequently favoring a robust representation of the frequently heard word. Learning these contrastive aspects during encoding may help newborns to form a more detailed representation of the familiarization word. In fact the recognition pattern in this experiment resembles that of Experiment 2 in which newborns were tested with words that differed from the familiarization word in only some phonemes. In Experiment 2, like in the current experiment, newborns probably relied on distinctive features of the sound to recognize the familiar word, which elicited a recognition response that involved the right frontal, left temporal and right parietal areas of the cortex.

Therefore, these findings support the proposal that variability of the input, rather than being detrimental, provide infants with valuable information for word-sound representations (Singh, 2008; Rost & McMurray, 2009, 2010). In the current study, we introduced variability by presenting two different words during familiarization, future studies using different tokens (e.g. a familiarization word pronounced by different speakers) will be able to address the question of whether newborns may also take advantage of such variability to create a more robust representation of words.

In this experiment we also found a significant decrement of oxy-Hb during the familiarization. This decrement was found in the left frontal area, which has been previously associated with encoding processes in numerous adult studies (Nyberg, Cabeza, & Tulving, 1996; Fletcher & Henson, 2001; Habib, Nyberg, & Tulving, 2003). This pattern of results suggests that the left frontal area could be actively involved in encoding processes from birth onwards. Presumably no such decrement was observed in any of our previous experiments, because encoding a single word in silent environments takes less time than encoding the same word in "noisy" environments. In this experiment, the decrement in the response is apparent within the first three blocks of the familiarization phase (see Figure 3.14). In the previous experiments, it is possible that the decrement of the response (associated with encoding) happened even within the first block of the familiarization phase. Thus, because the analysis is carried out over blocks, the actual decrement of the response may have been hindered. We did not attempted to explore word-related responses, thus this possibility remains to be explored in the future.

Finally, the present study shows that in certain situations intervening words might be beneficial for the encoding the –most frequent– familiarization word. It is an open question whether newborns are also able to create a representation of the less frequent sound (i.e. the intervening word), whether similar to the visual storage system, the newborns' auditory memory can retain only one item at the time (for instance see Káldy & Leslie, 2005 in the visual domain) or whether the auditory storage system can store two or more traces simultaneously. The memory span in auditory memory represents, thus, another avenue for future investigation.

3.5 Conclusion of the Chapter

A series of four experiments have been conducted to evaluate the mechanisms that induce newborns to forget words in conditions that resemble their natural surroundings. Our findings show that immediately after the learning phase, the traces are more susceptible to interference; the presentation of other speech sounds in the first minutes after familiarization produces retroactive interference (Experiment 4).

Yet, the last two experiments in this chapter showed that, even in the presence of other speech sounds, newborns' forgetting could be prevented or minimized. Experiment 6 showed that similar to adults, newborns benefit from distributed episodes of words to form a robust representation of the familiarization word. Furthermore, the results of Experiment 5 showed that the consolidation period is critical for the newborn brain to turn the sensory experience into an enduring memory trace. In particular, the consolidation that takes place in the first few minutes after the familiarization phase makes the trace more resistant to retroactive interference.

Additionally, the results show that when newborns listen to music during the consolidation period, the trace of the familiarization word remains intact (Experiment 3), suggesting that the interference effect that induces forgetting in newborns is selective to speech sounds. These results evidence that at birth humans' representational system is sensitive to the properties that differentiate speech and music stimuli. This initial distinction of sounds in the auditory storage system may be useful for newborns to start learning from speech.

We have shown that the first stages of language acquisition depend on many properties of memory that are functional since birth. Certainly the long-term storage of speech items will be also influenced by maturational changes of the brain and the mechanisms of consolidation, perhaps also by the emergence of native phonological categories, and the links between auditory and visual stimuli. If so, what does the representation of speech sequences look like in later stages of development? In the next chapter we will explore some mechanisms of memory that could influence word learning in 7 months-old infants.

Chapter 4

Vocabulary under construction: Features of short-term memory influence early word learning

Lexical acquisition starts much before infants utter their first words (Swingley, 2009; Gervain & Mehler, 2010). In the second semester of life speech production is still very limited, yet infants are able to recognize the sound of their own name, distinguishing it from similarly sounding names (Mandel, Jusczyk, & Pisoni, 1995); they can associate words like *mommy* and *daddy* with the respective faces of their parents (Tincoff & Jusczyk, 1999), they can recognize common words (body parts or food names) and associate them with pictures depicting these nouns (Bergelson & Swingley, 2012). Moreover, young infants can recognize words from storybooks (Jusczyk & Hohne, 1997), and they can establish object-word mappings in laboratory tasks (Friedrich & Friederici, 2011).

What does allow infants to begin the process of lexical acquisition? This is a question for which there is not a single answer. For instance, the fundamental capacities to remember words spoken in the surroundings and to associate the word sounds with their referents develop around the same age in which infants begin to attune to the phonetic categories of their native language. In fact, at the beginning of the second semester of life the discrimination of native speech contrasts improves (Patricia K. Kuhl, 2004; Patricia K. Kuhl et al., 2006) while non-native contrasts become difficult to detect (Werker & Tees, 1984; Werker & Tees, 1999), presumably refining the representation of word sounds and making them more stable in memory. Moreover, about the same age infants are able to segment potential word forms out of the continuous speech stream they hear, profiting from many different cues of the input such as transitional probabilities between adjacent syllables (Saffran, Aslin, & Newport, 1996), word-level stress patterns (Cutler, 1994), phonotactic regularities (Mattys, Jusczyk, Luce, & Morgan, 1999; Mattys & Jusczyk, 2000), allophonic variation (Jusczyk, Hohne, & Bauman, 1999), prosody (Shukla, Nespor, & Mehler, 2007; Shukla, White, & Aslin, 2011), and cues from other familiar sounds (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005).

Apart from the identification of the sounds that correspond to words, lexical learning should depend on the memory systems that sustain the retention of the newly heard sounds. In fact, the transition from short-term to durable representations of speech constitutes a key component of lexical acquisition, as suggested by a number of studies in healthy adults (Papagno, Valentine, & Baddeley, 1991; Papagno & Vallar, 1992), children (Gathercole & Baddeley, 1989; Gathercole, Hitch, Service, & Martin, 1997), neuropsychological patients (Baddeley, Papagno, & Vallar, 1988) and special developmental populations (Gathercole, Briscoe, Thorn, Tiffany, & ALSPAC, 2008) (see also Baddeley, Gathercole, & Papagno, 1998; Gathercole, 2006 for reviews). Even though the features of short-term memory should be particularly important for understanding the early stages of word learning, they have not been subject of systematic investigation in very young infants.

In this Chapter we explore some features of auditory short-term memory that could influence the encoding of speech in early infancy. Given that memory systems changes considerably during maturation (Nelson, 1995), and that the representation of speech sounds changes throughout development (Kuhl, 2004) it is possible that the features of memory that influence young infants' retention of verbal items do not resemble those in adults. On the other hand, certain properties of auditory memory might be intrinsic to the language-learning device; therefore they might be apparent at any stage in development, including early infancy (Baddeley et al., 1998).

For instance in adults, it is well established that when trying to repeat a sequence of words - or any other list of items- in the correct order, the first and the last items of the sequence (edges) are usually better remembered, whereas the items positioned in the middle of the sequence are more difficult to recall. This phenomenon is called the serial position effect (*hereafter SPE*) and it is one of the most robust effects among those observed in verbal memory tasks in adults (Crowder & Greene, 2000). The SPE became apparent since the first investigations on memory by Hermann Ebbinghaus, and has been repeatedly replicated in numerous studies in which memory for order is required. Moreover, SPE has been recently related to the adult's performance on artificial-grammar-learning experiments. Indeed, adults are able to generalize regular properties of a short speech sequence only when the regularities are positioned at the edges; in contrast, participants cannot learn these generalizations when the same regularities occurred in middle positions of words (Endress, Scholl, & Mehler, 2005; Endress & Mehler, 2009, 2010). The authors of these studies have argued that certain primitive perceptual and memory mechanisms might account for why certain portions of the speech signal (e.g. edge located items) are more readily learnable than others (Endress, Nespor, & Mehler, 2009). Would these primitives guide language learning also during development?

In the current studies we investigate word-learning capacities in 7 months-old infants by exploring whether general mechanisms of short-term memory, similar to those observed in adults, affect vocabulary learning during infancy. In particular, we focus on whether infants' short-term memory privilege the items located at the edge positions of a new multisyllabic item, namely whether infants, like adults, experience serial position effects.

We test the SPE in word learning because, like any other sequential learning tasks, learning a –multisyllabic– word requires remembering both the contents and the order of the elements that compose the sequence. In fact, the SPE is observed in adults not only when they have to learn a list of words but also when they are learning a multisyllabic item (Gupta, 2005; Gupta, Lipinski, Abbs, & Lin, 2005). If the readiness to learn items located at the edges is a consequence of primitive features of short-term memory, it might also influence infants' acquisition of words.

It should be noticed, however, that the paradigms commonly used to test SPE need to be adapted for preverbal infants. The traditional tasks often require adults and children to recall previously heard items; whereas in infants, recognition of the correct sequences of syllables -rather than explicit recollection- ought to be tested.

In the paradigm implemented in the current studies (adapted from Kovács & Mehler, 2009), a pentasyllabic word is paired with a puppet during the familiarization phase; in the test, participants hear words that differ from the familiarization word in either the edge syllables or in the middle syllables. Recognition will be measured through the infant's looking behavior when searching for the puppet after the presentation of each test word. Therefore the task resembles classical short-term memory tests in that it requires participants to remember the identity of the syllables, and also the order in which the different syllables appeared. Based on the adult studies we predict that the manipulations over the first and last syllables would affect infants' recognition more than changes in the internal syllables of the word.

Investigating word-learning abilities in this population offers an additional advantage for the study of proper short-term memory. Adults' performance on short-term memory tasks is influenced by previous semantic knowledge and other long-term associations (Hulme, Maughan, & Brown, 1991), instead seven month-olds possess a small vocabulary repertoire, which reduces considerably their possibilities to assimilate the new word by forming associations with previous experiences. In the following experiments we explore some properties of short-term memory in young infants hoping that these investigations could contribute to our understanding of the development of phonological memory in developmental populations.

4.1 Experiment 7: Serial order effects in early word learning

In this experiment we ask whether the ability of 7-month-olds to recognize familiar words is modulated by general properties of short-term memory, and if so, how. In particular we explored the serial position effect, a well know phenomenon observed in adults when learning sequences. We hypothesized that, if 7 month-olds experience the serial position effects when learning a pentasyllabic word, they should remember syllables at the edge positions better than syllables at the internal positions of the word.

4.1.1 Participants

Participants were twenty-two infants (12 females); aged from 6 months 15 days to 7 months 27 days (mean age 7 months 5 days). All participants were full term, with no birth complications, and no hearing or visual problems reported. Additional 4 infants were excluded from the analysis because of crying or fussiness (n=2), or because there were technical problems (n=2). Parents of the infants signed the informed consent after they had understood the procedure and after all their questions had been answered. The Ethics Committee of Scuola Internazionale Superiore di Studi Avanzati approved the study.

4.1.2 Stimuli

Linguistic stimuli were pentasyllabic CVCVCVCVCV words. Thirty-one CV syllables were constructed by pairing one vowel ([a], [e], [i], [o], [u]) and one consonant ([t], [m], [n], [d], [s], [v], [k], [g], [r], [l], [p], [b], [z], [f]). The C-V pairs served to generate different words that were used during the Familiarization as non-targets (see table 4.1). The syllables “so”, “tu”, “ma”, “ve”, “fi” were used exclusively to generate the two target words (i.e. *sotumavefi* presented to half of the infants; and *tusomafive* presented to the other half). Moreover, we made sure that the syllables “zi”, “bu”, “ne”, “lo” never appeared during the familiarization. Each word –either target or non-target– contained 5 different syllables constructed with five different consonants and five different vowels. All the words were nonsense in Italian (see table 4.2). The stimuli were synthesized with the female voice of the MBROLA Italian database IT4 (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996). The duration of each syllable was 200ms and a monotonous pitch of 240Hz was used. There were no pauses between syllables.

The visual stimuli consisted of one central attractor and puppets presented inside two white squares (one presented at the time). The two white squares had a side-length of 8cm, with a distance between them of 13.5 cm. The colorful puppets were used as visual reinforcement. In the Familiarization, one of the puppets was paired with the target word, whereas another two puppets were randomly paired with the non-target words. The pairing of the target word with the puppet was counterbalanced across participants.

FAMILIARIZATION PHASE
Non-targets
lumorisape
ganidemovu
zunikesano
banigukemo
puderimola
rudesakeni
todegusazi
vogukerita
bogusaderi
fagunimobe
vorikegusa
merizanigu
vusadenibo
tisamokezu
lukerideno
pesariguto

Table 4.1 Pseudowords used in the familiarization phase of Experiments 7, 8 and 9

FAMILIARIZATION PHASE	TEST PHASE			
Targets		**Experiment 7**	**Experiment 8**	**Experiment 9**
sotumavefi	Condition 1	sovematufi	sovematufi	sovematufi
	Condition 2	fitumaveso	sotumavefi	sozimalofi
or				
tusomafive	Condition 1	tufimasove	tufimasove	tufimasove
	Condition 2	vesomafitu	tusomafive	tunemabuve

Table 4.2 Target words used in the familiarization and test phases of Experiments 7, 8 and 9.

4.1.3 Procedure

An adaptation of the Kovács & Mehler (2009) procedure, implemented previously in word-learning tasks for young infants (Hochmann et al., 2011) was used in this study. The experiment consisted of a familiarization phase of 32 trials and a test phase of 8 trials (see figure 4.1).

Trials in the familiarization phase started with a display of a central visual attractor and two white squares, one on the left and one on the right side of the screen. When the infant fixated the central attractor, one word was played. After the offset of the word, the central attractor disappeared, leaving only the two white squares visible for 1s. Then a looming puppet appeared on one of the two white squares for two seconds. The puppet was accompanied by a tinkling bell of 300ms presented with a delay of 800ms with respect to the onset of the visual stimulus. The target word (either *sotumavefi* or *tusomafive*) was presented in sixteen familiarization trials. This word predicted the location of the puppet in one side of the screen. The location was counterbalanced across participants. Other sixteen words were used as non-targets; these words preceded the puppet's appearance in the other side of the screen. Target and non-target trials were presented in an interspersed pseudo-random order. We ensured that targets or non-targets were presented at the most three times in a row.

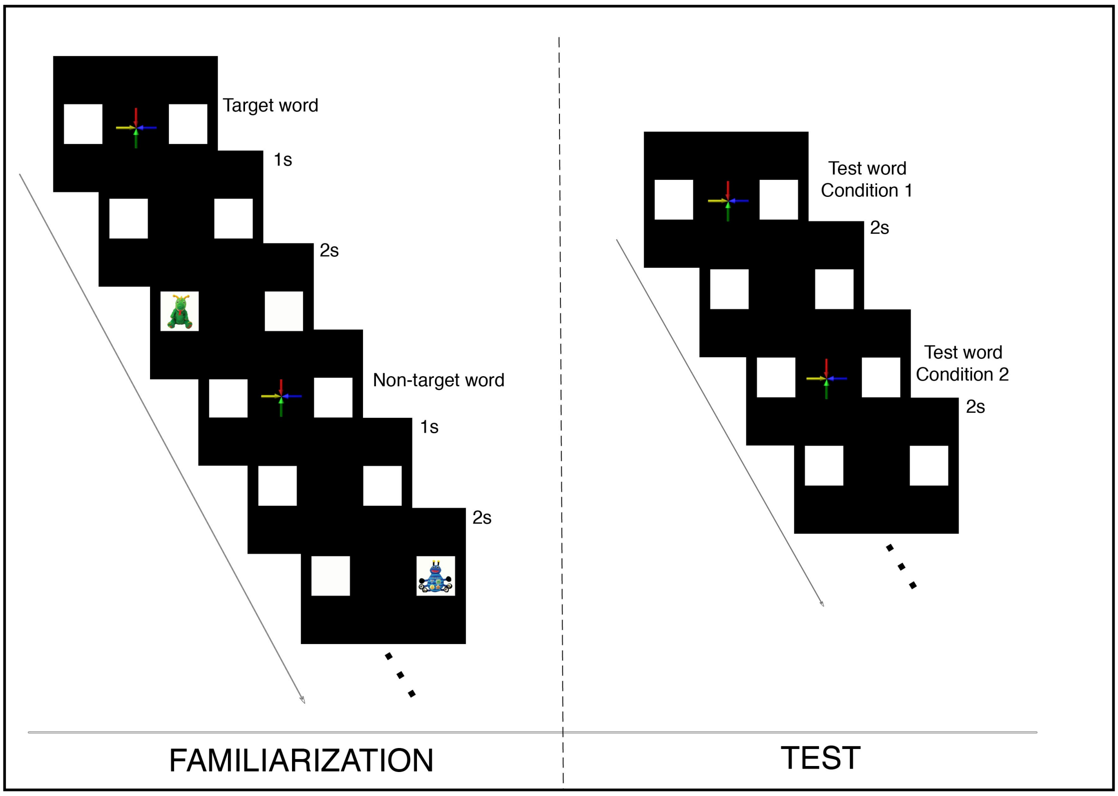

Figure 4.1 Schematic diagram of the procedure used in Experiments 7, 8 and 9.
The experiments consisted of a familiarization phase (32 trials) and a test phase (8 trials). In the familiarization phase, there were 16 target trials and 16 non-target trials randomly interleaved. Trials started with a fixation display showing a central visual attractor. In each trial, participants listened to one word (either target or non-target). After the offset of the word, the anticipatory period (duration, 1s) began. At the end of the anticipatory period a puppet appeared for 2s on one side of the screen. Afterwards, the next trial began. In the test phase no puppet appeared after each test word.

The test trials were similar to the familiarization trials, except that after each word no puppet was displayed in either side of the screen. Two kinds of test words were constructed with the same five syllables that constituted the target word in the familiarization phase. In this way each test word differed from the target word only in the serial order of the syllables. In both conditions the position of two syllables was interchanged. However, in one case we switched the position of the first and the last syllables (Edge-switch condition), whereas in the other condition the second and fourth syllables were switched (Internal-switch condition). For instance, if the target word in the familiarization phase was *sotumavefi*, the test words were *fitumaveso* (Edge-switch condition) and *sovematufi* (Internal-switch condition). Each test word was presented four times in the test in a pseudo-random order. The first four trials consisted in two trials of each condition. Two seconds after the word offset, the next trial started.

4.1.4 Data acquisition

Infants' gaze was recorded with an eye-tracker (TOBII 1750). The eye tracker was integrated into a 17-inch TFT screen, where the stimuli were presented via an IMAC 10,1 running PsyScope X software (http://psy.ck.sissa.it/). A loudspeaker was placed behind the screen for the presentation of the acoustic stimuli. Infants were seated on their parent's lap at about 50 cm distance from the monitor. A hidden videocamera was used to observe the infant's behavior.

4.1.5 Data analysis

We measured the infant's first fixation[6] after hearing the test word and before the beginning of the next trial. We divided the screen into three equal parts: left, middle

---

[6] Looking times of 80 ms or more were considered a fixation.

and right. Infants were coded as looking either to the target side or to the non-target side (left or right sides of the screen).

We computed normalized difference scores by subtracting the number of trials in which the infant first fixated the target and non-target sides, divided by the total number of trials in which the infants fixated either side: (#trials fixating the target side - # trials fixating the non-target side)/ (#trials fixating the target side + # trials fixating the non-target side). The difference scores were compared to chance level (zero) using two-tailed one-sample t-tests. Positive difference scores indicate that infants searched for the puppet on the target side (where the puppet appeared after the target word during the familiarization phase), whereas negative difference scores indicate that infants searched for the puppet on the non-target side (where the puppet appeared after the non-target words during the familiarization).

4.1.6 Results

Familiarization phase

A linear regression analysis computed over the proportion of infants displaying correct anticipatory looks, showed no tendency for more correct anticipations in later familiarization trials for either target ($y=0.0035x+0.459$, $R^2=0.020$) or non-targets ($y=-0.0067x +0.5518$, $R^2=0.094$). During the familiarization phase infants anticipated 53% of the target trials with 50% accuracy. In the non-target trials they anticipated 55% of the time; they were correct in 48% of the trials.

Test phase

The results over the first fixation data are presented in Figure 4.2. Infants looked significantly above chance to the target side in the Internal-switch condition (mean difference score =0.33; t(21) = 2.123; *p* < 0.05; Cohen's d= 0.45), and they were at chance for the Edge-switch condition (mean difference score =0.19; t(21) = 1.156; *p* = 0.262; Cohen's d= 0.26) [7].

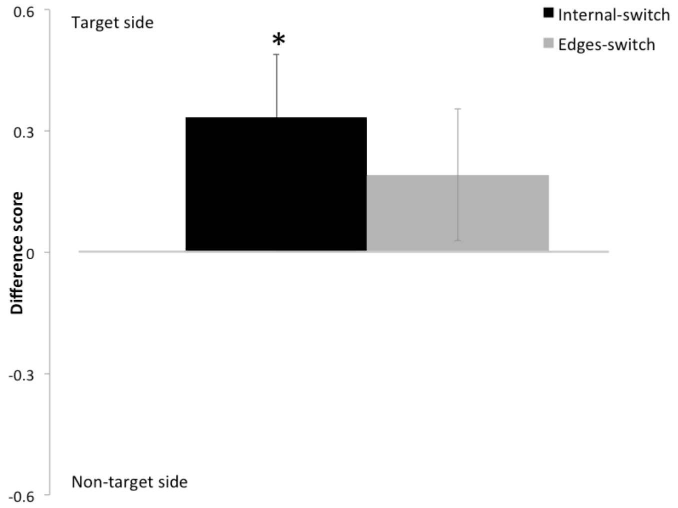

Figure 4.2 Experiment 7: Normalized difference scores considering first fixations.
The y-axis shows the mean difference scores. Positive difference scores indicate that infants searched for the puppet on the target side while negative difference scores indicate that infants searched for the puppet on the non-target side. Bars depict standard error of the mean (SEM). * p<0.05.

---

[7] Using longest fixation and overall accuracy as dependent variable we obtained similar results. Longest fixation: Internal-switch condition mean difference score= 0.36 [t(21) = 2.792; *p* = 0.01; Cohen's d= 0.595]; Edge-switch condition mean difference score = -0.075 [t(21) = -0.439; *p* = 0.665; Cohen's d = -0.09]. Overall accuracy: Internal-switch condition mean difference score = 0.23 [t(21) = 1.81; *p* = 0.08; Cohen's d= 0.397]; Edge-switch condition mean difference score = 0.15 [t(21) = 1.34; *p* = 0.19; Cohen's d= 0.29].

4.1.7 Discussion

We found that the recognition of the syllables in a multisyllabic word was directly related to their position in the sequence. Participants recognized the familiar sequence only when the edge-syllables were kept in place and even though the middle syllables were swapped. However, infants failed to recognize the sequence when the first and the last syllables were swapped and the remaining syllables were kept in the original positions. Thus, the results correspond to the predictions based on the SPE. Infants encoded better the syllables at the boundaries of the sequence, whereas the syllables in the middle positions were encoded much less reliably.

The logic used to turn the familiar word into the test word was the same in both conditions (i.e. the order of the familiar sequence was altered by interchanging the position of two syllables), yet, the test words were not equivalent in terms of the transitional probabilities. In the Internal-Switch condition all possible transitional probabilities between the syllables of the familiar word were altered, so that no syllable followed the same syllable that was predicted by the familiar sequence. In the Edges-Switch condition only two out of four transitional probabilities were disrupted. Thus, if infants in this task had used the transitional probabilities between adjacent syllables to track the familiar sequence, they should have showed better recognition for the Edges-Switch condition than for the Internal-Switch condition. Instead, infants failed to recognize the word in the Edges-Switch condition and succeeded in the Internal-Switch condition. This indicates that participants did not rely on the mutual information of syllables to identify the familiar sequence. The results rather suggest that, similar to children and adults, 7-month-olds are more influenced by general short-term memory features when learning a multisyllabic sequence. They memorize better the syllables at the edges of the sequence than the middle syllables. Notably, these results cannot be accounted for by statistical analysis of first order alone[8]. Saffran *et al.* (1996) demonstrated that statistical information of the input plays an important role during lexical learning. Specifically, the authors showed that 8-month-old infants are able to use transitional probabilities (i.e., the probability with which one syllable predicts the next or the previous one) to segment continuous speech into its constituent words. In the current study, each word was presented in isolation, thus no segmentation task was required to identify the sound of the word. However this should have not prevented the infant from tracking the statistical information between the syllables.

Another interpretation of the results obtained in the current study is that the paradigm itself predisposed infants to choose only one of the test sequences. Perhaps participants recognized the familiar syllables in both test items, but maybe they had a greater difficulty to associate two different words with the same puppet. Indeed, previous studies on interleaving learning of regularities (Kovács, 2008; Kovács & Mehler, 2009) or interleaving word learning (Hochmann et al., 2011), have shown that due to limitations in their executive functions, most monolingual infants tend to learn and associate only one item or family of items with a referent (i.e. side of the screen), even though two words or regularities had been provided.

In the current experiment infants were exposed only to one word-object association during the familiarization, however, during the test they heard two very similar sounds. Thus, infants' apparent recognition of only one sequence might have been the

---

[8] Participants could have also focused on the non-adjacent dependencies between the first and last syllables of the words. The statistical mechanism triggered by non-adjacent dependencies might be useful for the generalization of structural information (Peña, Bonatti, Nespor, & Mehler, 2002; Endress & Bonatti, 2007; Marchetto, 2009)

consequence of limitations in their ability to associate two sounds with the correct puppet (or side of the screen). We will explore this possibility in Experiment 8. Infants will be presented with two test words that can be potentially associated with the puppet. We will test participants in the Internal-switch condition (in which the second and fourth syllables are interchanged) and we will also present them with the Target condition (presenting the same word of the familiarization phase). Based on the results of experiment 7, we expect that infants recognize the Internal-switch word and associate it to the side of the screen where the puppet appeared. We also expect infants to recognize the Target word in the test phase, because they have learned this word and its association with the puppet throughout the familiarization phase of the experiment. If infants are able to associate two similar words with the corresponding object/side of the screen, they should perform above chance in the two conditions, otherwise they should perform above chance only in one of them.

4.2 Experiment 8: Serial position effects? A control study

In the previous experiment we have showed that 7 month-olds experience serial position effects when encoding a multisyllabic word. In fact, infants only recognized the sequence that maintained the edge syllables in the original positions. However, above we have also considered the possibility that the nature of the paradigm used in experiment 7 may have prevented infants from displaying recognition of two words. We now seek to rule out this possibility (while simultaneously replicating our primary result) by investigating whether infants can recognize two words in the test whose edge syllables were the same syllables of the familiarization word.

4.2.1 Participants

Participants were twenty infants (11 males) aged from 7 months 0 days to 8 months 6 days (mean age 7 months 19 days). All participants were full term infants, with no birth complications, and no hearing or visual problems reported. Additional infants were excluded from the analysis because of crying or fussiness (n=5), because they looked to the screen only in one test trial (n=2), or because there were technical problems (n=1).

4.2.2 Stimuli

The auditory and visual stimuli were identical to those used in Experiment 7.

4.2.3 Procedure

We studied participants while they were learning a pentasyllabic nonsense word. The word was paired with a specific puppet/location in the screen. The apparatus and the procedure were the same as in Experiment 7. In the test, we measured the infant's looking behavior when searching for the puppet. In one condition, the test word differed from the familiarization word in the position of two syllables: the second and the fourth syllable were switched (Internal-switch condition). In the other condition, the test word was identical to the target word learned during the familiarization phase (Target condition). For instance, if the target word in the familiarization phase was *sotumavefi*, the test words were *sotumavefi* (Target condition) and *sovematufi* (Internal-switch condition) (see table 4.2). No puppet ever appeared during the test phase. There were four trials per condition. Trials were presented in a pseudo-random order: the first four trials consisted in two trials of each condition. Two seconds after the word offset, the next trial started.

4.2.4 Data Analysis

For the scoring and data analysis we measured infants' first fixations as described in Experiment 7.

4.2.5 Results

Familiarization phase

A linear regression analysis computed over the proportion of infants displaying correct anticipatory looks, showed no tendency for more correct anticipations in later familiarization trials (target trials $y=0.0010x+0.5769$, $R^2=0.001$; non-target trials $y=0.0090x+0.3114$, $R^2=0.166$). Infants anticipated in 71% of the target trials. From these, they anticipated correctly in 57% of the trials, and incorrectly in 43% of the trials. Participants also anticipated in 71% of the non-target trials, however they were correct only in 44% of the trials.

Test phase

Considering the first fixation data we found that, as in Experiment 7, infants looked significantly above chance to the target side in the Internal-switch condition (mean difference score =0.33; $t(19) = 2.297$; $p < 0.05$; Cohen's d = 0.51). Moreover infants looked significantly above chance to the target side also in the Target condition (mean difference score =0.36; t(19) = 2.462; $p < 0.05$; Cohen’s d = 0.55) (see figure 4.3).

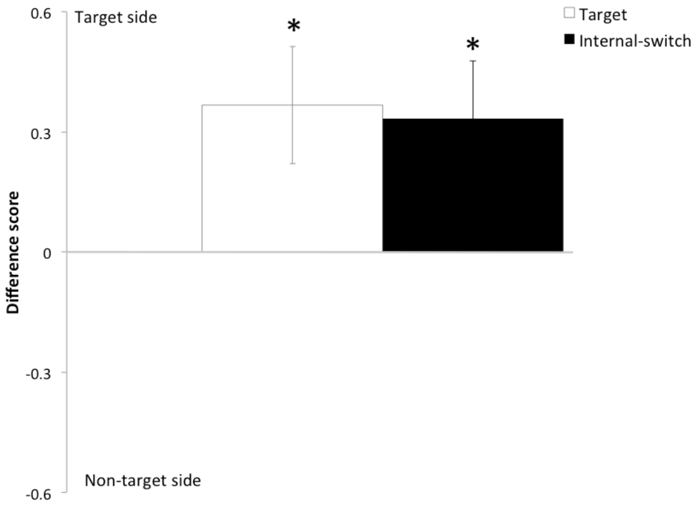

Figure 4.3 Experiment 8: Normalized difference scores considering first fixations.
Infants searched for the puppet on the target side both in the Target Condition and in the Internal-Switch Condition. The y-axis shows the mean difference scores. Bars depict standard error of the mean (SEM). * p<0.05.

4.2.6 Discussion

The current experiment replicates the main result from Experiments 7: participants searched for the puppet in the target side (the side associated to the target word in the familiarization phase) when the syllables of the edges remained in the original positions. Given that infants showed a similar looking behavior in the Internal-switch Condition and in the Target Condition, we rule out the possibility that the results of experiment 7 were due to infants’ failure to associate two test words with the corresponding side of the screen. The data rather confirm the initial interpretation:

speech representation in early word-learning tasks is modulated by fundamental short-term memory properties, similar to those classically observed in adults and children.

The first and the last items of a sequence may thus constitute cornerstones of the initial phonological representations that infants store while learning words. In fact, we observed that preserving the first and last syllables at the original positions in the multisyllabic word is necessary for the infants' recognition of the word.

If words are being represented mainly by the edges, the question that remains is whether infants also encode the identity of the middle syllables being unable to remember their exact sequential order or whether because of memory limitations they are disregarding completely the middle syllables of the word. These two processes – remembering the occurrence of an item, and remembering where in a sequence that item occurred – might be independent (Henson, 1998). In fact, whereas the brain structures that mediate the representation of temporal order (Fortin, Agster, & Eichenbaum, 2002) may not be functionally mature until the end of the 1st year of life (Nadel & Zola-Morgan, 1984; Nelson, 1995), the recognition of previously heard sounds, seems to be available since birth (see chapters 2 and 3).

Experiments 7 and 8 indicate that infants do not track precise information about the internal syllables of the pentasyllabic word. In the following experiment we ask whether these results arise because infants cannot encode any information about the internal syllables, or just the information about the sequential order. To explore this possibility we will test infants with items that differ from the familiarization word either in the order or in the identity of the middle syllables. If participants cannot encode the sounds of the middle syllables at all, any change in the internal syllables – either item change or sequence change– should go unnoticed. On the other hand, if infants are able to encode the sound of all syllables but not their correct sequential order, then participants should differentiate the target word from a word that contains completely different syllables.

4.3 Experiment 9: A developmental dissociation between item recognition and serial order recognition

Learning a word requires remembering both the identity and the order of the syllables that compose it. In the previous experiments we showed that when learning a pentasyllabic word, participants remember better the syllables at the edge positions as compared to syllables at the internal positions. Does this mean that 7 month-olds are unable to encode any information about the internal syllables? In this experiment we investigated this issue by asking whether participants can recognize a word that differs from the familiarization word in the identity of the middle syllables.

4.3.1 Participants

Twenty infants (12 males; mean age = 7 months 15 days; age range = 6 months 20 days to 8 months 6 days) were included in the analysis. All participants were full term infants, with no birth complications, and no hearing or visual problems reported. Additional infants were excluded from the analysis because of crying or fussiness (n=3), because they looked to the screen only in one test trial (n=1), or because of equipment failure (n=3).

4.3.2 Stimuli

The stimuli were the same of previous experiments, except that in the test phase four syllables: “zi”, “bu”, “ne”, “lo” were used to replace the second and fourth syllables in the Internal-change condition (see Procedure in the next section). After this manipulation the test words -as the familiarization words- contained 5 different syllables with no adjacent repetition of vowels neither of consonants (See table 4.2). All the words had no meaning in Italian. The stimuli were synthesized with the female voice of the MBROLA Italian database IT4 (Dutoit et al., 1996). The duration of each syllable was 200ms and a monotonous pitch of 240Hz was used.

4.3.3 Procedure

The procedure was identical to the previous experiments. In one test condition the test word differed from the familiarization word in the position of two syllables: the second and the fourth syllable were switched (Internal-switch condition). In the other condition, the test word also differed from the familiarization word in the second and fourth syllables, although not in the position of the syllables but in their identity (Internal-change condition). For instance, if the familiarization word was *sotumavefi*, the test words were *sovematufi* (Internal-switch condition) and *sozimalofi* (Internal-change condition). Each condition was presented four times in the test in a pseudo-random order. The first four trials consisted in two trials for each condition. Two seconds after the word offset, the next trial started.

4.3.4 Data Analysis

As in the previous experiments we measured infants’ first fixations.

4.3.5 Results

Familiarization phase

A linear regression analysis computed over the proportion of infants displaying correct anticipatory looks, showed no tendency for more correct anticipations in later familiarization trials (target trials $y=0.012x+0.3426$, $R^2=0.106$; non-target trials $y=0.004x+0.4061$, $R^2=0.020$). Infants anticipated in 60% of the target trials. From these, they anticipated correctly in 52% of the trials, and incorrectly in 48% of the trials. Participants anticipated in 55% of the non-target trials and they were correct in 44% of them.

Test phase

Consistently with the experiments reported above, infants looked significantly above chance to the target side in the Internal-switch condition (mean difference score =0.38; $t(19) = 2.364$; $p < 0.05$; Cohen's d= 0.54). In contrast, participants did not perform significantly above chance in the Internal-change condition (mean difference score =0.075; $t(19) = 0.462$; $p = 0.649$; Cohen's d= 0.103) (see figure 4.4).

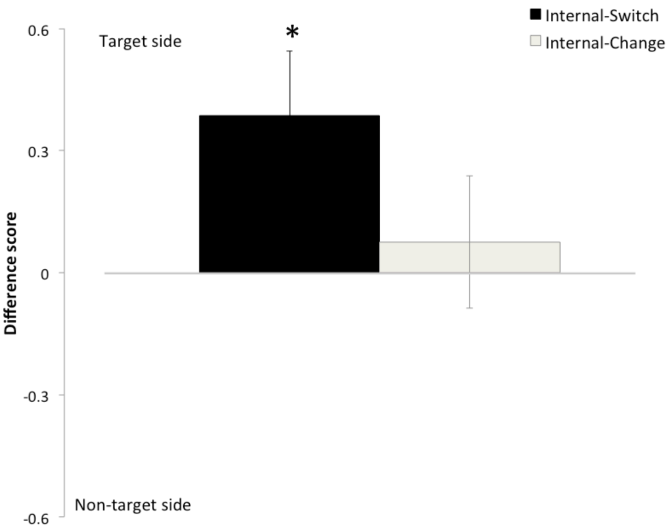

Figure 4.4 Experiment 9: Normalized difference scores considering first fixations.
Infants searched for the puppet on the target side only in the Internal-Switch Condition. The y-axis shows the mean difference scores. Bars depict standard error of the mean (SEM). * $p<0.05$.

4.3.6 Discussion

The results show that infants looked significantly above chance for the puppet in the side of the screen previously associated to the target word, only when the middle syllables swapped positions but not when these syllables were substituted by new ones. This suggests that infants are able to encode some information about the identity of the middle syllables but they have difficulties in accurately encoding the sequential order. In both conditions the edge syllables remained in the original positions, but only in one of these conditions infants recognized the sequence. It thus seems that the syllables of the edges are a necessary requisite, yet not a sufficient condition for the recognition of the familiar sequence.

To be able to learn the sound of a specific word, infants need to remember the contents of the syllables that compose the word and the sequence in which they are presented. A series of three experiments have been conducted to evaluate the phonological short-term memory abilities of 7-month-old infants when learning new words. In our studies participants were first familiarized with a pentasyllabic word associated with a toy appearing at a specific location of the screen (left or right). Subsequently in the test, we found that infants searched for the puppet when listening to the familiarization word (indicating that they learned the association between the word and the referent) and also when listening to a word whose first and last syllables were the same of the familiarization word –while the second and fourth syllables were interchanged–. In contrast, when listening to a word whose first and last syllables were switched, infants did not search for the puppet at the corresponding side of the screen, confirming the relevance of keeping the syllables at the edges to license word recognition.

Therefore, congruent with the hypothesis that features of short-term memory influence infants' first representations of words, the data show that at 7 months of age infants encode the syllables at the edges more accurately than the syllables in the middle position of the word, in other words, that infants like adults experience the serial position effect.

Explanations of how adults maintain serial order in short-term memory are a subject of debate (Burgess & Hitch, 1999; Henson, 1998; Page & Norris, 1998; Brown, Preece, & Hulme, 2000). Our data comply with the view that the elements located at the beginning and the end of sequences are more salient, and as such they are represented more accurately than the internal items that are encoded relative to the edges (Henson, 1998; 1999).

One might argue that our results, rather than reflecting serial position effects, are a consequence of the infants' impossibility to encode a long multisyllabic word; forcing infants to remember only the first phoneme or syllable of the sequence. In fact, this possibility would be enough to explain why the replacement of the first syllable, in the test phase of experiments 7 and 8, prevented participants from recognizing the familiarization word. The results of experiment 9, however, call for a different explanation. The data evidenced that the identity of the internal syllables was also relevant for the recognition of the familiarization word. Infants were able to identify the specific set of syllables that form the pentasyllabic word; they failed to recognize, however, the order in which the internal syllables were organized. Our results are in agreement with a number of studies showing the same phenomenon in adults, namely that participants of serial learning tasks retrieve the correct list of items, but they often retrieve them in the incorrect order (e.g. Bjork & Healy, 1974; Healy, 1974; Henson, Hartley, Burgess, Hitch, & Flude, 2003). In fact, the assumption that order and item information are two separate components is often included in the models of memory for serial order information (Henson, 1998; Burgess & Hitch, 1999; Brown et al., 2000). In addition to confirm that there are two separate components, our data suggest that these two components follow different developmental pathways: recognition of the identity of elements seems to emerge early in life, whereas a detail recognition of serial order would develop later in development, probably with the maturation of the brain structures that mediate the representation of temporal order (Nelson, 1995; Fortin et al., 2002).

Even though the serial position effect had been previously observed in infants when performing visual recognition tasks (Cornell & Bergstrom, 1983; but see also Gulya, Rovee-Collier, Galluccio, & Wilk, 1998 in the motor domain), showing SPE during speech processing has particular interest for understanding the first stages of vocabulary and language learning. In fact, being able to encode the elements located at the edges might be useful for infants when acquiring most linguistic regularities that appeal to positions of items, for example, when learning the position of the lexical stress. Stress is assigned in either word-initial (as in Hungarian), word-final (as in Turkish), or in another syllable that is counted from one of the word edges (as in Italian, where stress generally falls on the second syllable from the last); but no language assigns stress on positions that are not defined relative to the edges (Halle & Vergnaud, 1990; Hayes, 1995; Kager, 1995). The saliency of edges seems to be also informative to learn various grammatical regularities of language. For instance, prefixes and suffixes are cross-linguistically much more frequent than infixes (Greenberg, 1957; Slobin, 1973 as cited in Endress et al., 2009) and reduplications are found mainly at edges of morphological constituents (Broselow & McCarthy, 1983).

Further evidence in support of the view that edges play an important role for extracting relevant linguistic information come from artificial grammar experiments in adults. In the experiments participants noticed structural regularities located at the edges of sequences, but had greater difficulty when the regularities were in the middle of sequences (Endress et al., 2005, 2009; Endress & Mehler, 2009, 2010). Thus, in addition to its potential role for early lexical acquisition, the results from the artificial grammar experiments suggest that edges detection might help storing information useful for rule generalization.

Yet another example of the crucial role of edges during language acquisition comes from the study of Seidl & Johnson (2006). In the study, the authors compared 8-month-olds' recognition of words appearing in utterance-initial, utterance-medial, or utterance-final position. Participants recognized words that appeared initially or finally in the passages, however infants did not show recognition of words from the middle of the utterances, suggesting that placing words at the beginning or at the end of utterances or phrases increases the likelihood of infants recognizing the word. Moreover, it has been observed that mothers tend to place words they are teaching their infants at the end of sentences (Aslin, Woodward, LaMendola, & Bever, 1996), a spontaneous –probably inadvertent– strategy that could be effective to transmit prominent information to infants.

The acquisition and maintenance of new language information, is a critical human ability that is needed throughout the life span. Like children and adults, very young infants experience SPE when encountering for the first time a sequence of nonsense verbal items, suggesting that the same properties of short-term memory modulate language acquisition from early infancy onward. However, the current results also suggest that the temporary storage that sustains early language acquisition is independent of the ability to –actively– rehearse words. While this remark may seem somewhat trivial when referring to our results in preverbal infants –who cannot pronounce words–, it contrasts with influential theoretical models of short-term memory in adults (e.g. Baddeley & Hitch, 1974; Baddeley, 2003). Based on a great number of studies (e.g. Murray, 1968; Baddeley, Thomson, & Buchanan, 1975; Cowan et al., 1992), such models propose that a rehearsal –articulatory based– process maintains the traces of speech in working memory (the multicomponent that mediate language learning), however they do not make predictions about the way preverbal infants would maintain new phonological representations in short-term memory, when for instance, learning words.

A possible implication of the current study is that the mechanisms used to maintain the phonological information in infants do not necessarily rely on articulation. According to this assumption, only later in development, when infants start pronouncing words, the temporary storage of verbal information would be influenced by cognitive operations such as rehearsal. The alternative to this assumption would be that internal speech-motor programs are in place before infants utter their first words (see for instance Kuhl & Meltzoff, 1982; 1984; 1996; Bristow et al., 2009), and that such links between auditory and motor representations give rise to mechanisms in preverbal infants that resemble rehearsal in adults. The acceptance of one or the other alternative might be premature at present, thus the exact way preverbal infants maintain the traces of the phonological representation when recruited for working memory tasks (for instance word learning) awaits clarification.

4.4 Conclusion of the chapter

Altogether the studies of the present Chapter show that, when learning a new pentasyllabic word, 7 month-old infants experience serial position effects. These results are consistent with previous research on sequential memory showing that learners encode better the elements at the boundaries whereas the items in the middle positions are encoded much less reliably. Moreover, the results correspond to previous views that suggest a separation between storage of the content and storage of the order of elements; in addition our data showed a possible developmental dissociation between these two storage systems.

We argue that the relative advantage of syllables located at the edges of a word should not be considered a limitation for lexical learning, but rather another feature of human memory that could be incorporated in the current views of language acquisition. The elements at the edges seem to be learned particularly easily, this general feature of learning could interact directly with the language acquisition device, explaining why crucial information is spontaneously transmitted to the infants at the edge of the sentences and why edge-based regularities are frequently found in human languages.

Chapter 5

General Discussion

Language is a special faculty of our species, and memory is one of the prerequisites to learn language. Even though verbal memory has been extensively studied in adults (e.g. Ebbinghaus, 1964; Baddeley & Hitch, 1974), few studies had been devoted to explore it in very young infants, leaving unresolved questions related to the actual influence of memory over the language acquisition path, and the similarities and differences between early and mature speech sounds representations. The primary aim of this thesis was precisely to investigate the origins of memory for speech in the first stages of development. In particular, the studies explored features of the memory systems that support word-sound retention, and certain properties of these systems that prevent the long-term storage of sounds in newborns and young infants.

The studies showed that the initial state of auditory memory in humans is influenced by the species' predisposition to learn language, more specifically that the format with which word sounds are stored in the newborn's representational system might be relevant only to linguistic stimuli. At the same time, the studies showed that domain-general features of memory constrain the early stages of language development. It is argued that this bidirectional interaction between memory capacities and early language abilities provides crucial information for understanding language development. The characterization of memory for speech sounds in newborns and young infants also served to identify certain properties of memory that are used to construct the conjectural model on memory development with which we conclude this thesis.

Each of the empirical findings that support the above-mentioned model will be unfolded in the following sections. Chapters 2, 3 and 4 will be reviewed separately, highlighting the features of early memory systems that derive from the findings. The discussion also includes possible implications for the theories of language acquisition and suggests some directions for future research.

5.1 Pre-phonological memory: the format of speech representations at birth

What does the sound of a word look like in the representational system of a newborn? Is a new speech sound represented in the same fashion in the memory system of a newborn and that of an adult person? In Chapter 2 we focused on the elements that newborns remember from the sound of words, and on identifying the brain structures supporting the recognition of these elements. Two fNIRS studies provided information regarding memory for speech stimuli in newborns, revealing some similarities and differences in the means that adults and young infants possess to acquire language.

In Experiment 1 we found evidence that infants possess a representational system that allow them to remember word sounds since birth. We will call this representational system the pre-phonological memory system (see figure 6.1) because it stores speech information before native phonological categories have been established. It is assumed that the establishment of phonological categories later in development, would sustain more stable word representations.

A feature of the pre-phonological memory that can be derived from the results of Experiment 1 is precisely the format of a word trace. Because the familiar and the novel words with which newborns were tested had the same syllabic structure, stress pattern, duration, intensity, and were recorded by the same speaker, we concluded that the contents of the trace must transcend these general acoustic features. The system might have encoded at least some spectral features of the sound; otherwise the newborn's brain would have not been able to distinguish between the novel and familiar sounds.

We also concluded that the newborn's representation of a word in the pre-phonological system should be more stable than a trace in sensory memory. In fact, the pre-phonological system retained the trace of the word for a 2-minute interval, whereas sensory traces generally decay in about 1,5sec in the newborns' brains (Cheour, Čėponienė, et al., 2002). Auditory sensory memory sustains the newborn's well-attested abilities to discriminate sounds. Instead, we suggest that the pre-phonological system would sustain a trace that, because it lasts longer than a sensory memory, would be available to trigger specific learning mechanisms.

Apart from revealing its non-purely sensory/acoustic nature, few other conclusions could be obtained from the first experiment concerning the characteristics of the proposed pre-phonological memory system. Despite of the fact that the familiar and the novel sounds were controlled for general acoustic features, in the first experiment these words contained different sets of phonemes. This left unresolved the questions regarding the specific elements that the system stores from words and whether the format of a word sound at birth resembles the format of the representation of words in adults. Moreover, the distributed activation pattern in the cortex did not allow us to identify specific brain structures responsible for the recognition response. Considering the substantial phonemic differences between the two words, it seemed clear that this wide cortical activation might had reflected the involvement of other cognitive and perceptual processes besides memory. However we reserved our conclusion in this regard, until more information was gathered in the following experiments.

In the second experiment of the series, we thus focused on investigating whether newborns can recognize the specific elements that compose a word, and whether the pre-phonological representation are similar to the adults' word representations. To do that, we tested newborns with words that shared with the familiarization word only some phonemes (instead of testing participants with the same or totally different phonemes, as we did in the previous experiment). We chose test words that shared with the familiarization word either the vocalic-tier or the consonantal-tier. This choice was made because previous theoretical and experimental work had shown that adults (Van Ooijen, 1996; Lee et al., 2001; Cutler et al., 2000; Nespor et al., 2003; Bonatti et al., 2005; Toro, Nespor, et al., 2008; Carreiras & Price, 2008) and older infants (Nazzi, 2005; Havy & Nazzi, 2009; Nazzi et al., 2009; Hochmann et al., 2011) represent consonants better than vowels when processing, remembering and learning words. Therefore, the contrast between these two categories enabled a first comparison between early and mature representations.

The results showed that newborns, unlike adults and older infants, encode better the vowels of a word sound, or at least recognize them more readily as compared to the consonants. Interestingly, as discussed in Chapter 2, the prominence of vowels in the pre-phonological representations cannot be explained in terms of discriminability. Newborns' possess extraordinary discrimination capacities for consonantal and for vocalic contrasts (e.g. Eimas et al., 1971; Eimas & Miller, 1980; Bertoncini et al., 1987; Cheour-Luhtanen et al., 1995; Dehaene-Lambertz & Pena, 2001; Kushnerenko et al., 2001), which suggests that the differences between consonants and vowels might directly arise at the representational level in the pre-phonological system.

What would the primary purpose of this representational difference be? Could it reflect core aspects of the language acquisition device and its interaction with memory? If so, how could this distinction modulate the first stages of language acquisition?

It is possible that the very selection of some (but not all) elements of the speech signal ensures a more efficient process of language acquisition. In fact, some experimental theoretical views (Nespor et al., 2003) suggest that there is a division of labor between consonants and vowels, such that one speech category preferentially supports the acquisition of the lexicon (consonants), whereas the other one (vowels) is favored for extracting and generalizing structural properties. This theory argues that the complementary roles of consonants and vowels allow the language acquisition device to work simultaneously on the extraction of regularities and on the acquisition of lexical items (see also Hochmann, 2010; Hochmann, Benavides-Varela, Nespor, & Mehler, 2011). This might suggest that the basic distinction between consonants and vowels must be embedded in the pre-phonological system because it enables the young infant to learn the language more efficiently.

Still, newborns stored mainly the vowels of speech sequences, which, at least in adults and older infants, favor the mechanisms to process and identify, not lexical items but syntactic structures (Toro, Nespor, et al., 2008; Toro, Shukla, et al., 2008; Pons & Toro, 2010). This suggested that, even though we presented participants with isolated words, newborns' brains might not be prepared to efficiently represent words (i.e. speech units), but rather to focus on the structural properties cued by prosodic information (Morgan & Demuth, 1996; Nespor & Vogel, 2007).

This conclusion complies with the views that prosodic information (which is mainly carried by vowels) may help neonates to acquire the first glimpse of knowledge about their native language, guiding them to recognize the rhythm (Mehler et al., 1996) and to "bootstrap" into structural properties during development (Morgan & Demuth, 1996). This interpretation could be additionally justified by the localization of the recognition response. In this experiment the recognition response involved areas of the brain that are frequently associated with language and memory processes (left temporal and right parietal areas), however the most pronounced effects were found in the right frontal area of the newborn's brain, which is regularly activated in adults and infants during processing of prosodic information (Friederici & Alter, 2004; Homae et al., 2006). Yet, the right frontal area is also involved in retrieval processes of verbal material in adults (Shallice et al., 1994; Rugg et al., 1996; Fletcher & Henson, 2001), which may also suggest that basic mechanisms of recognition memory that involve this area of the brain are already active at birth.

Moreover, the fact that the representational system uses mainly vowels to store word sounds indicates that the system undergoes a qualitative change during the first year of life. In fact, as stated before, by the end of the first year of life infants start attributing separate roles to vowels and to consonants (Hochmann et al., 2011).

Twelve-month-olds prefer vowels to extract syntactic structures, but they mainly use consonants to establish the phonological representation of words (instead of relying on vowels as newborns do). In Chapter 2 we conjectured about possible causes of this developmental transition. We discussed the role of maturational changes, exposure to speech, and computational prerequisites for lexical memory. We emphasized that the emergence of the lexical role of consonants may be triggered by the establishment of native consonantal categories about the second semester of life. When long-term representations (lexical entries) need to be established, the memory system may represent mainly the consonants because they are more stable and are perceived more categorically than vowels (see also Hochmann, 2010; Hochmann et al., 2011). Before these native consonantal categories are formed, the system may rely on the information carried by vowels, even to recognize the sound of words.

Does it imply that the newborns' representation of speech is less or more accurate than an adult representation? It is premature to answer such question relying only on the information we have gathered until now, however it is an issue that deserves further investigation. In particular it would be interesting in the future to investigate whether these developmental differences in encoding are related to the easiness with which young infants engage in language acquisition. Since the seminal work of Lenneberg (1967) it is known that there is a biologically determined window of opportunity in which language can be acquired natively. He proposed that once this critical period is over, language-learning abilities degrade, and the resulting acquisition will not be native-like. Since then, a large body of evidence has accumulated, clearly establishing that infants are better first and second language learners than adults (e.g. Newport, 1990), however it is not entirely clear what the basis of this advantage is. The answer might be in part related to this developmental transition of the representational system.

To summarize, two properties of the pre-phonological system were derived from the results obtained in Chapter 2. Experiment 1 showed that the system is more than sensory/acoustic in nature, and Experiment 2 showed that the system represents mainly the information carried by vowels (Experiment 2). These properties might indicate that the initial state of the human's auditory memory has some features that cannot be generalized outside the language domain. Other properties of the pre-phonological system that further support this argument will be discussed in the following section.

5.2 Pre-phonological memory: potentials and constraints

Understanding language acquisition would not be complete if the factors that limit the learning process in everyday situations were not identified. In fact, it is not clearly understood why, for instance, infants display recognition of highly familiar words (i.e. their own name) only at about four or five months of age (Mandel *et al.,* 1995), given that many perceptual and (some) memory mechanisms necessary for language processing are already present in newborns. In Chapter 3 we performed a series of experiments to explore some properties of the pre-phonological storage system in situations that approximate better the newborns' common environment. In particular, we focused on investigating the factors that prevent this memory system from retaining the word-traces in "noisy" situations.

Experiment 4 showed that the presence of other speech sounds during the retention interval can disrupt a word-trace that has been recently stored in the pre-phonological system. This suggested that a strong interference *speech-to-speech* could induce forgetting in young infants, preventing the lexical acquisition process to begin from the first months of life. However, additional studies provided a finer understanding of the interference phenomena in newborns by showing that in certain circumstances the word-trace can be maintained in the pre-phonological system despite the presence of other speech sounds. In fact, in Experiments 5 and 6 we were able to show that it is possible to promote an enduring representation of the familiar trace such that the same amount and type of intervening sounds that produced interference in Experiment 4, could not affect the traces in these studies.

To turn the familiarization trace into a representation that could resist interference, we manipulated the structure of the encoding and retention phases of the original experiment. In particular, Experiment 6 showed that the familiarization word endures the interfering experience when it is learned in distributed –as opposed to massive– episodes. In Chapter 3 we discussed a number of reasons why this organization of the learning phase could stabilize the representation of the familiarization word. The benefits of distributed learning might stem from the variability of the input. The interspersed presentation of the “to-be-remembered” word and another word could have highlighted the contrastive aspects of the words, which in turn facilitated the encoding of more detailed elements of the word. The pattern of activation in this experiment supported this interpretation. In fact the recognition response involved the same brain areas that were activated in Experiment 2, in which newborns had to recognize specific elements of the words in order to find the distinctions between the familiar and test words.

In Experiment 5 we found another condition in which the traces of the familiarization word can resist the presence of interfering sounds. Newborns' brains distinguished the familiar and the novel words, despite of the fact that the same amount and type of interfering stimuli were presented during the retention interval. Crucially, in this experiment the interfering words were presented two minutes after the familiarization phase, whereas in Experiment 4 they were presented immediately after familiarization.

From the point of view of the consolidation theory, this result has a straightforward interpretation: the manipulation allowed the traces of the word to begin the consolidation process and consequently to become more resistant to interference. In fact, our results closely resemble the effects observed in adult and animal studies in which new stimuli regularly produce stronger interference when they are presented shortly after previous learning than when they are presented long time afterwards (e.g. McGaugh, 2000 for a review).

In Experiment 5, unlike previous experiments, we found a recognition response confined to the right frontal area (previously we had observed more distributed responses). This pattern of activation suggested that the pre-phonological representation had undergone some qualitative change in these experimental conditions. At least three possible explanations could account for this change. First, the localized response could have been caused by the fact that the trace began to decay with time (the familiarization and the test phases of this experiment were separated by 4 minutes whereas in other experiments they were separated by a 2-minute interval). Second, the consolidated trace may have become weaker after the newborns heard the interfering words. Thus, in contrast with Experiment 4, in which the trace was completely lost after interference, in this experiment the interfering sounds may have only diminished the strength of the trace. Third, the very process of consolidation could have caused the shrinking of the network. Different studies suggest that consolidation is associated with a reorganization of the brain response; for instance declarative memories, that are originally dependent of the hippocampus, become independent of this region of the brain after consolidation (Dudai, 2004). Thus, in a similar way consolidation processes in this study may have modified the pattern of response leading to the activation of a more limited area in the cerebral cortex. In other words, consolidation may have transformed the initial representation into a representation that became independent of some brain areas that had originally sustained the information storage.

If the first or the second alternatives described above were correct, they would challenge the actual contribution of this early consolidation mechanism for the maintenance of the traces in everyday situations. They would imply that even when the machinery that triggers the consolidation process is present in newborns, it might not be sufficient to maintain a lasting representation of the word because after 4 minutes and/or the presence of interfering sounds, the trace grows smaller. The third alternative, however implies the opposite: that the reduction of the cortical network is related to the strengthening of the trace rather than to its fading. Experiment 5 is not conclusive in terms of the contribution of this early consolidation mechanism for the construction of lasting speech representations. The exact cause of this localized response needs to be clarified in future studies before any definitive interpretation can be provided. Still, a clear conclusion regarding the pre-phonological system can be derived from experiments 5 and 6. These experiments showed that pre-phonological representations are susceptible to certain factors that promote a resistant memory trace.

In addition to the endurance of the word traces in the presence of speech sounds, we also found that newborns are able to remember the familiarization word in the presence of intervening music. More specifically, the results of Experiment 3 showed that, unlike speech, music does not interfere with the traces of the familiarization word, even if presented immediately after the familiarization phase. In fact, the recognition response elicited a distributed cortical network similar to the one observed in Experiment 1, in which the familiarization and test phases were separated by silence. This suggests that the pre-phonological storage maintains the properties of the trace in a similar way in silent conditions and in the presence of music.

In Chapter 3 we extensively discussed the selectivity of the representational system and considered possible accounts for this phenomenon. We discussed some perceptual and cognitive aspects that could yield different types of representations for words and for music stimuli. In particular, we were concerned with aspects such as the acoustic differences between speech and music, the differences in the sources that produced the two types of stimuli and the evidence that vocal and non-vocal sounds have different neural substrates in adults (Belin et al., 2000, 2002; but see also Grossmann, Oberecker, Koch, & Friederici, 2010; Belin & Grosbras, 2010). We also discussed the results of Experiment 3 in light of the innate disposition to attend to biologically produced stimuli (e.g. speech) as compared to non-biologically produced sounds (e.g. Peña et al., 2003; Vouloumanos & Werker, 2004). Finally we discussed the possibility that speech and music were processed by two different mental modules and therefore relied on two separate representations in the newborns' brain (Fodor, 1983; Peretz & Coltheart, 2003).

Independently of what the ultimate explanation of this intriguing effect turns out to be, it is clear that the pre-phonological system is sensitive to the properties that distinguish speech and other non-speech sounds such as music sequences. Whereas in this study we were able to observe this general differentiation within the auditory domain, future studies manipulating different properties of the familiar and intervening sounds will be able to single out the specific properties of the signal that give rise to such separation. For instance, systematically manipulating the proximity between the speech sequences and music excerpts (e.g. presenting a humming lullaby instead of the instrumental melody during the retention interval) may provide important information about how newborn infants represent speech, and which properties of speech and music (either acoustic or abstract properties) are most relevant for the separate representations we have observed.

Altogether, the results of Chapter 3 suggest that the pre-phonological system is able to maintain the immediate speech experience in a format that resists interference from music. In certain circumstances the system also has the potential to retain word sounds in the presence of other speech sounds. This is enabled by processes that can act upon the trace during encoding (Experiment 6) or after learning (Experiment 5). The most important constraint of the pre-phonological system is that, in the absence of these enhancing processes, it is susceptible to *speech-to-speech* interference that disrupts the trace.

Besides the interplay between these memory processes, there might be linguistic factors that also constrain the establishment of long-term representation in the first months of life, for instance, the lack of native phonological categories. The emergence of the phonological categories around the second semester of life provides a more stable representation of the sound of words. This helps the infant to map the sound into a more defined acoustic space and consequently to recognize this sound in a more efficient way. The emergence of the phonological categories represents a crucial step in language acquisition, and as discussed in Chapter 2, might be at the basis of the change in the format of the representation of words. The presence of these categories may be conceived as a prerequisite for the existence of a phonological memory as it has been traditionally studied in adults (e.g. Baddeley, 2003). Yet, the emergence of the phonological categories does not prevent the language acquisition device to be influenced by other properties of memory, as it will be illustrated below in the discussion of Chapter 4.

5.3 Serial position effects: a feature of memory determining early word learning

Lexical learning requires the construction of lasting memory representations for word forms, word meanings, and for referential connections between these representations. In order to figure out the appropriate links of the new sound with the available referents, word learning heavily relies on the infants' ability to maintain new phonological information in the storage system while "reviewing" new and previously learned word-meaning associations, an ability that seems to emerge at the end of the second semester of life (see Chapter 4).

Experiments 6 and 7 showed that at the beginning of the lexical acquisition process, and despite their potential capacity to store the segmental details of a new lexical entry, 7 month-old infants store more accurately the sounds that are located at the edges of a word than those located in the middle positions. Thus, infants experience serial position effects during the acquisition of lexicon. These results suggested that general mechanisms of memory continue affecting vocabulary learning throughout infancy.

We observed that when learning multisyllabic words infants may use the first and the last elements of the multisyllabic words as the anchors of the representations. However, the results of Experiment 9 showed that infants also encoded the phonological information of the internal syllables; they were not able, however, to track the precise sequential order of these syllables. This suggests not only that the two components (order and content) are dissociated in the infants' initial representations, but also that they may follow different developmental pathways. Whereas the identity of sounds can be stored earlier in development, the information about the order of sounds seems to emerge later in life. An extensive work needs to be done in order to determine the neurodevelopmental causes of this dissociation.

In Chapter 4 we comprehensively discussed different linguistic phenomena arguing that the relative advantage of syllables located at the edges of a word should not be considered a limitation for lexical learning, but rather another feature of human memory that needs to be incorporated in the current views of language acquisition. The prevalence of the elements located at the edges is likely to interact with the language acquisition device explaining, for instance, why crucial information is spontaneously transmitted to the infants at the edge of the sentences (Aslin et al., 1996) and why edge-based regularities are frequently found in most languages of the world (see Endress et al., 2009).

Chapter 6

General Conclusions

This thesis has investigated the initial state of memory for the sound of words. We explored the constraints and potentialities of the information storage system in newborns, and the properties of memory that sustain word learning in older infants. The results provide evidence for the existence of an auditory memory system that shapes language development, and at the same time is shaped by the human disposition to acquire language.

We conclude that the course of language development depends not only on linguistic operations, but also on certain features of this representational system. The studies also show that the initial state of memory displays some species-specific properties, in spite of the fact that memory has always been considered a general cognitive capacity that is not specific to language.

Based on the results of this thesis, we propose that newborns are endowed with memory capacities to retain speech sequences in a semi-specialized storage system that we called a *pre-phonological system* (see figure 6.1). The studies of this thesis consistently showed that the pre-phonological system activate the right frontal area of the newborns' brain, that is homologous to areas involved in memory retrieval in adults. At least five properties of this system were identified in our studies. The first property concerns its nature. The information storage system that assists short-term speech representations is more than purely sensory in nature. In fact, it stores enough information from speech as to distinguish between acoustically similar words (Experiment 1), and the time span is longer than sensory memory.

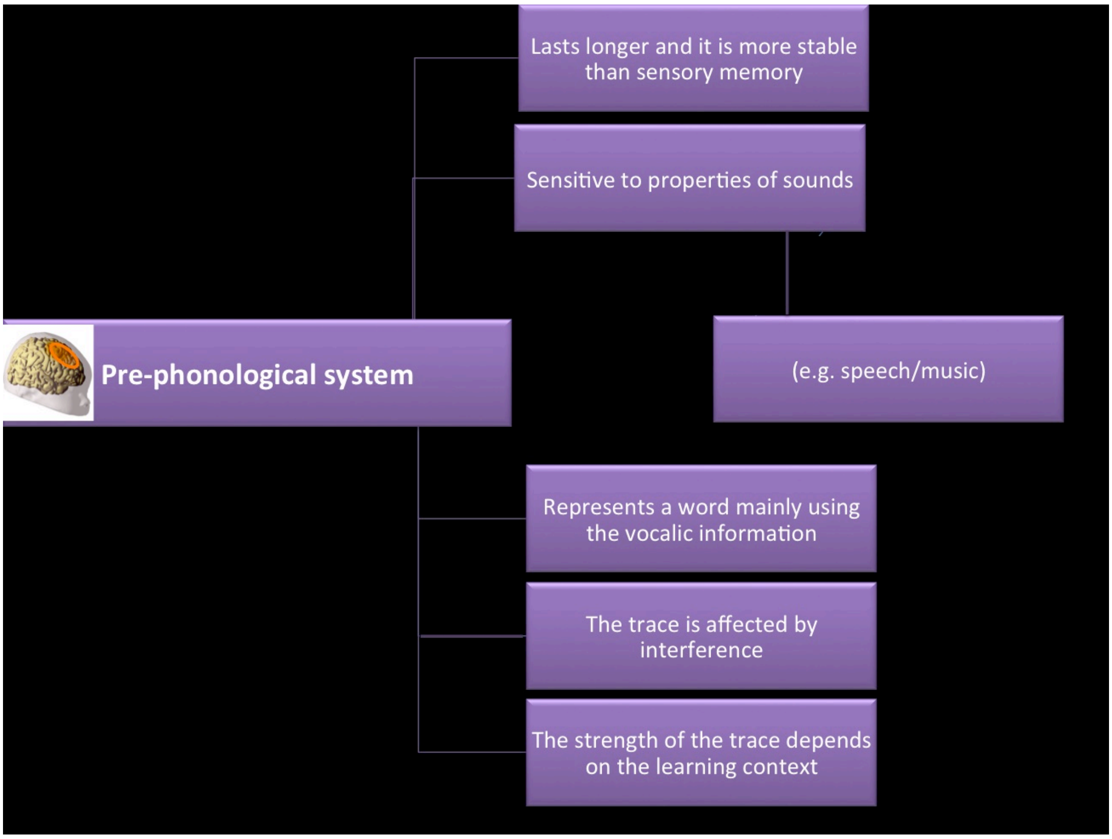

Figure 6.1 A description of the properties of the pre-phonological system in newborns: a hypothetical model of the earliest word representations.

The second property concerns its specificity. It was shown that this pre-phonological system represents speech sequences in a format that does not generalize to non-speech sounds such as instrumental music (Experiment 3). We argue that this basic division within the representational system may derive from a species-specific bias that helps neonates to engage in extracting information from speech. The third property of the system concerns the amount of differentiation within the pre-phonological storage. We observed that not every speech sound is represented to the same extent in the newborn's brain. Instead, in the newborns' representation of speech sequences there is a basic distinction between the speech categories (consonants and vowels) or at least over the information that these speech categories convey (Experiment 2). This early distinction between consonants and vowels may inform the learner about distinctive and complementary aspects of language -lexicon and syntax- (Nespor et al., 2003; Hochmann et al., 2011).

Other properties of the pre-phonological system may not be specific to language, but may be determinant of its ontogenesis. For example, the interference effects (Experiment 4) might be one of the causes of the delayed engagement in lexical acquisition observed in young infants.

It was also shown that the traces of a word sound could be strengthened through distributed learning (Experiment 6) and consolidation (Experiment 7), therefore enduring in the presence of interfering sounds. The development of the neural structures underlying the consolidation process constitutes, therefore, another important factor to take into account in the current views of language development.

It is likely that other language-specific variables such as the development of native phonemic categories also contribute to form stable representations of word sounds in the memory. If so, the system required for lexical learning in 7 month-olds may already differ from the pre-phonological memory system proposed for the maintenance of verbal information in newborns, both in its format and its functionality. It differs in its format because the sounds would be presumably stored in a more stable way, due to the emergence of the phonemic categories. As a consequence, this memory system should provide more regular cues for word identification than those supplied by the pre-phonological memory system in newborns. Moreover, this system should differ from the pre-phonological storage in its function because older infants store speech sequences while learning the association of a novel sound with a referent, unlike newborns who presumably cannot yet establish such associations; the retention of the sound and its meaning may require a more robust storage system.

We also conceive the representational system of preverbal infants (of about 7 months of age) as differing from verbal memory models in adults that propose an articulatory-based mechanism to sustain the traces of speech. The underlying assumption is that such models cannot be applied to preverbal infants who (most probably) do not rehearse words using their articulatory systems. The integration of auditory and motor representation should have important implications for the representational system because it provides a greater number of potential cues for recognition (or retrieval).

Given these differences, we conceive the existence of an intermediate state between the pre-phonological system and the memory systems in adults and children. Two properties of this intermediate representational system can be derived from our studies: the first and the last elements of a multisyllabic word serve as the anchors of these representations. Moreover, as discussed earlier, two components (order and content) are dissociated in the representations in this intermediate system. Extensive work needs to be done in order to identify additional properties of this system and to further explore similarities and differences in the way preverbal infants and adults store speech information while learning new words.

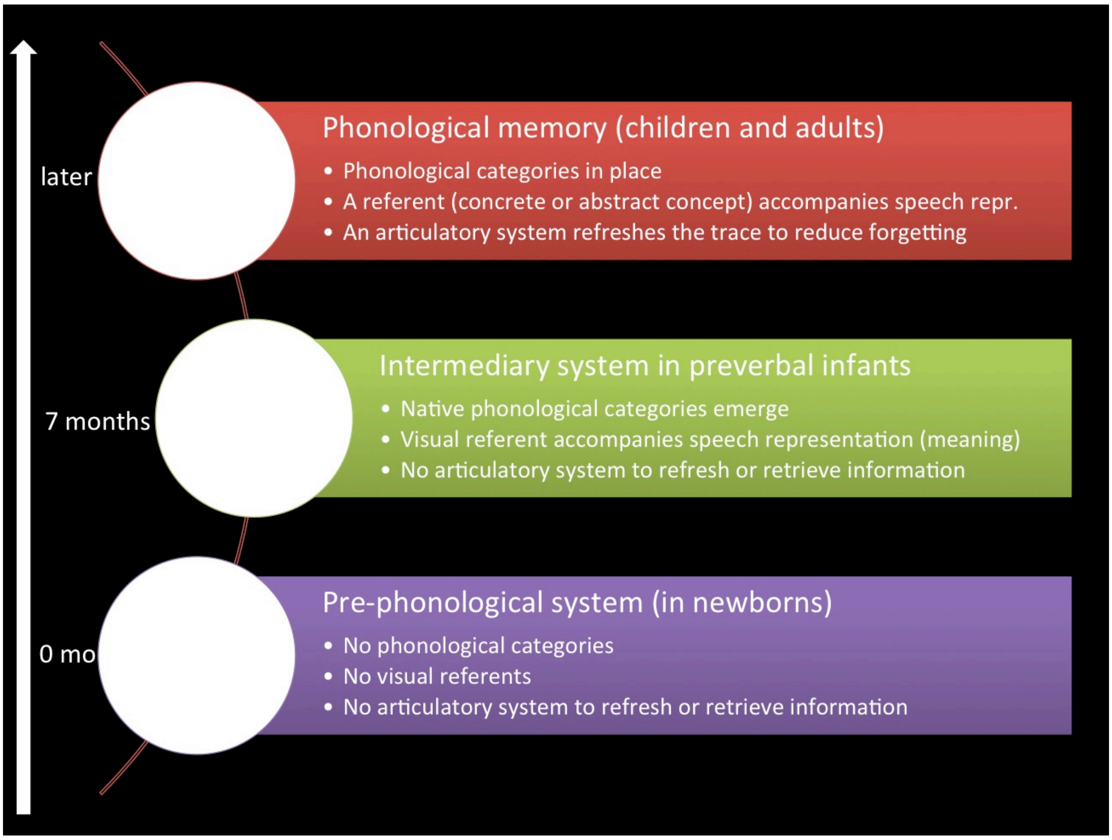

Figure 6.2 Timeline of the hypothetical changes of speech representation in short-term memory.

What arises from this thesis is a view of the auditory memory system as a semi-specialized cognitive capacity at the service of the language acquisition device since birth. The same memory system can also be seen as a general cognitive capacity (not specific to the linguistic domain) that, however, has the power of modulating the language acquisition process. As this representational system facilitates and/or constrains the learning that is based on speech (the most important sound for our species) we maintain that it is fundamental to better understand its properties in the different stages of development. We hope that this thesis could serve as a valid endeavor for this venture.

References

Aldridge, M. A., Stillman, R. D., & Bower, T. G. R. (2001). Newborn categorization of vowel-like sounds. *Developmental Science*, *4*(2), 220-232. doi:10.1111/1467-7687.00167

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: a simple network model. *Proceedings of the National Academy of Sciences*, *91*(15), 7041 -7045.

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415-445. doi:10.1016/j.jml.2003.08.006

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063-1087. doi:10.1037/0278-7393.20.5.1063

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. *Memory.*, Handbook of perception and cognition (2nd ed.). (pp. 237-313). San Diego, CA, US: Academic Press.

Aslin, R. N., & Mehler, J. (2005). Near-infrared spectroscopy for functional studies of brain activity in human infants: promise, prospects, and challenges. *Journal of Biomedical Optics*, *10*(1), 11009. doi:10.1117/1.1854672

Aslin, R. N., Woodward, J., LaMendola, N., & Bever, T. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in* *early acquisition* (pp. 117-134). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes (Vol. Volume 2, pp. 89-195). Academic Press.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat Rev Neurosci*, *4*(10), 829-839. doi:10.1038/nrn1201

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2009). *Memory* (1st ed.). Psychology Press.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158-173. doi:10.1037/0033-295X.105.1.158

Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, *27*(5), 586-595. doi:10.1016/0749-596X(88)90028-9

Baddeley, A.D., & Hitch, G. (1974). *Working Memory* (Vol. 8). New York, NY, US: Academic Press.

Baddeley, A.D., & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(2), 176-189. doi:10.1016/S0022-5371(70)80048-2

Baddeley, Alan D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575-589. doi:10.1016/S0022-5371(75)80045-4

Barnes, J. M., & Underwood, B. J. (1959). “Fate” of first-list associations in transfer theory. *Journal of Experimental Psychology*, *58*(2), 97-105. doi:10.1037/h0047507

Bauer, P. J. (2006). Constructing a past in infancy: a neuro-developmental account. *Trends in Cognitive Sciences*, *10*(4), 175-181. doi:10.1016/j.tics.2006.02.009

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129-135. doi:10.1016/j.tics.2004.01.008

Belin, P., & Grosbras, M.-H. (2010). Before Speech: Cerebral Voice Processing in Infants. *Neuron*, *65*(6), 733-735. doi:10.1016/j.neuron.2010.03.018

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17-26. doi:10.1016/S0926-6410(01)00084-2

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309-312. doi:10.1038/35002078

Benavides-Varela, S., Gómez, D. M., & Mehler, J. (2011). Studying Neonates' Language and Memory Capacities with Functional Near-Infrared Spectroscopy. *Frontiers in Psychology*, *2*. doi:10.3389/fpsyg.2011.00064

Benjamini, Y., & Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics*, *29*(4), 1165-1188.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1113380109

Bertoncini, J., Bijeljac-Babic, R., Blumstein, S., & Mehler, J. (1987). Discrimination in neonates of very short CVs. *The Journal of the Acoustical Society of America*, *82*(1), 31. doi:10.1121/1.395570

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, *117*(1), 21-33. doi:10.1037/0096-3445.117.1.21

Best, C. C., & McRoberts, G. W. (2003). Infant Perception of Non-Native Consonant Contrasts that Adults Assimilate in Different Ways. *Language and Speech*, *46*(2-3), 183 -216. doi:10.1177/00238309030460020701

Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior*, *13*(1), 80-97. doi:10.1016/S0022-5371(74)80033-2

Bliss, T. V. P., & Collingridge, G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature*, *361*(6407), 31-39. doi:10.1038/361031a0

Bliss, T. V. P., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of Physiology*, *232*(2), 331 -356.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci*, *8*(3), 389-395. doi:10.1038/nn1409

Bonatti, L., Frot, E., Zangl, R., & Mehler, J. (2002). The Human First Hypothesis: Identification of Conspecifics and Individuation of Objects in the Young Infant. *Cognitive Psychology*, *44*(4), 388-426. doi:10.1006/cogp.2002.0779

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic Constraints on Statistical Computations. *Psychological Science*, *16*(6), 451 -459. doi:10.1111/j.0956-7976.2005.01556.x

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2007). On Consonants, Vowels, Chickens, and Eggs. *Psychological Science*, *18*(10), 924 -925. doi:10.1111/j.1467-9280.2007.02002.x

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and Me. *Psychological Science*, *16*(4), 298 -304. doi:10.1111/j.0956-7976.2005.01531.x

Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature*, *382*(6588), 252-255. doi:10.1038/382252a0

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J.-F. (2009). Hearing Faces: How the Infant Brain Matches the Face It Sees with the Speech It Hears. *Journal of Cognitive Neuroscience*, *21*(5), 905-921. doi:10.1162/jocn.2009.21076

Broselow, E., & McCarthy, J. (1983). A theory of Internal Reduplication. *The Linguistic Review*, *3*(1), 25-88.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*, 127-181. doi:10.1037/0033-295X.107.1.127

Bunsey, M., & Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, *379*(6562), 255-257. doi:10.1038/379255a0

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551-581. doi:10.1037/0033-295X.106.3.551

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nat Rev Neurosci*, *9*(8), 613-625. doi:10.1038/nrn2459

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature*, *403*(6768), 428-430. doi:10.1038/35000206

Carreiras, M., & Price, C. J. (2008). Brain Activation for Consonants and Vowels. *Cerebral Cortex*, *18*(7), 1727 -1735. doi:10.1093/cercor/bhm202

Carreiras, M., Vergara, M., & Perea, M. (2009). ERP correlates of transposed-letter priming effects: the role of vowels versus consonants. *Psychophysiology*, *46*(1), 34-42. doi:10.1111/j.1469-8986.2008.00725.x

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis. *Psychological Bulletin*, *132*, 354-380. doi:10.1037/0033-2909.132.3.354

Cheour, M., Martynova, O., Naatanen, R., Erkkola, R., Sillanpaa, M., Kero, P., Raz, A., et al. (2002). Psychobiology: Speech sounds learned by sleeping newborns. *Nature*, *415*(6872), 599-600. doi:10.1038/415599b

Cheour, Marie, Čėponiené, R., Leppänen, P., Alho, K., Kujala, T., Renlund, M., Fellman, V., et al. (2002). The auditory sensory memory trace decays rapidlyin newborns. *Scandinavian Journal of Psychology*, *43*(1), 33-39. doi:10.1111/1467-9450.00266

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., et al. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, *82*(1), 53-58.

Chiron, C., Jambaque, I., Nabbout, R., Lounes, R., Syrota, A., & Dulac, O. (1997). The right brain hemisphere is dominant in human infants. *Brain: A Journal of Neurology*, *120 (Pt 6)*, 1057-1065.

Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of Prosodic Boundary Correlates by Newborn Infants. *Infancy*, *2*(3), 385-394. doi:10.1207/S15327078IN0203_6

Christophe, A., Nespor, M., Teresa Guasti, M., & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: the case of the head-direction parameter. *Developmental Science*, *6*(2), 211-220. doi:10.1111/1467-7687.00273

Conrad, R., & Hull, a. J. (1964). Information, Acoustic Confusion and Memory Span. *British Journal of Psychology*, *55*(4), 429-432. doi:10.1111/j.2044-8295.1964.tb00928.x

Cooke, S. F., & Bliss, T. V. P. (2006). Plasticity in the human central nervous system. *Brain*, *129*(7), 1659 -1673. doi:10.1093/brain/awl082

Cornell, E. H., & Bergstrom, L. I. (1983). Serial-position effects in infants' recognition memory. *Memory & Cognition*, *11*, 494-499. doi:10.3758/BF03196986

Cotelli, M., Abutalebi, J., Zorzi, M., & Cappa, S. F. (2003). Vowels in the buffer: a case study of acquired dysgraphia with selective vowel substitutions. *Cognitive Neuropsychology*, *20*(2), 99-114. doi:10.1080/02643290244000158

Courage, M., & Cowan, N. (2009). *The development of memory in infancy and childhood*. Taylor & Francis.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory.

Journal of Memory and Language, *31*(1), 1-17. doi:10.1016/0749-596X(92)90002-F

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684. doi:10.1016/S0022-5371(72)80001-X

Crowder, R. G. (1976). *Principles of learning and memory*. Oxford, England: Lawrence Erlbaum.

Crowder, R. G., & Greene, R., L. (2000). Serial learning: Cognition and behavior. *in Tulving, E., Craick F. (eds). The Oxford handbook of memory*. New York, NY, US: Oxford University Press.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*, *5*(6), 365-373. doi:10.3758/BF03210660

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua*, *92*(0), 81-104. doi:10.1016/0024-3841(94)90338-7

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition*, *28*(5), 746-755. doi:10.3758/BF03198409

DeCasper, A., & Spence, M. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, *9*(2), 133-150. doi:10.1016/0163-6383(86)90025-1

DeCasper, AJ, & Fifer, W. (1980). Of human bonding: newborns prefer their mothers' voices. *Science*, *208*(4448), 1174 -1176. doi:10.1126/science.7375928

Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allirol, L., Dubois, J., Hertz-Pannier, L., & Dehaene, S. (2010). Language or music, mother or Mozart?

Structural and environmental influences on infants' language networks. *Brain and Language*, *114*(2), 53-65. doi:10.1016/j.bandl.2009.09.003

Dehaene-Lambertz, G., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport*, *12*(14), 3155-3158. doi:10.1097/00001756-200110080-00034

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional Neuroimaging of Speech Perception in Infants. *Science*, *298*(5600), 2013 - 2015. doi:10.1126/science.1077066

Deutsch, D. (1970). Tones and Numbers: Specificity of Interference in Immediate Memory. *Science*, *168*(3939), 1604 -1605. doi:10.1126/science.168.3939.1604

Dudai, Y. (2004). The Neurobiology of Consolidations, Or, How Stable is the Engram? *Annual Review of Psychology*, *55*, 51-86. doi:10.1146/annurev.psych.55.090902.142050

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). The MBROLA project: towards a set of high quality speech synthesizers free of use for non commercial purposes. *, Fourth International Conference on Spoken Language, 1996. ICSLP 96. Proceedings* (Vol. 3, pp. 1393-1396 vol.3). Presented at the , Fourth International Conference on Spoken Language, 1996. ICSLP 96. Proceedings, IEEE. doi:10.1109/ICSLP.1996.607874

Ebbinghaus, H. (1964). *Memory: a contribution to experimental psychology.* (H. A. Ruger & C. E. Bussenius, Trans.)Über das Gedächtnis.English. New York: Dover Publications.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science*, *171*(3968), 303 -306. doi:10.1126/science.171.3968.303

Eimas, P., & Miller, J. (1980). Contextual effects in infant speech perception. *Science*, *209*(4461), 1140 -1141. doi:10.1126/science.7403875

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*(2), 247-299. doi:10.1016/j.cognition.2006.09.010

Endress, A. D., & Mehler, J. (2009). Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology*, *62*, 2187-2209. doi:10.1080/17470210902783646

Endress, A. D., & Mehler, J. (2010). Perceptual constraints in phonotactic learning. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 235-250. doi:10.1037/a0017164

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, *13*(8), 348-353. doi:10.1016/j.tics.2009.05.005

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The Role of Salience in the Extraction of Algebraic Rules. *Journal of Experimental Psychology: General*, *134*(3), 406-419. doi:10.1037/0096-3445.134.3.406

Ewan, W. G. (1975). Explaining the intrinsic pitch of vowels. *The Journal of the Acoustical Society of America*, *58*(S1), S40. doi:10.1121/1.2002115

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104-113. doi:37/0012-1649.20.1.104

Fletcher, P. C., & Henson, R. N. A. (2001). Frontal lobes and human memory. *Brain*, *124*(5), 849 -881. doi:10.1093/brain/124.5.849

Fodor, J. (1983). *The Modularity of Mind*. The MIT Press. Retrieved from http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0262560259

Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience*. doi:10.1038/nn834

Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*, *89*(2), 267-276. doi:16/S0093-934X(03)00351-1

Friedrich, M., & Friederici, A. D. (2011). Word learning in 6-month-olds: fast encoding-weak retention. *Journal of Cognitive Neuroscience*, *23*(11), 3228-3240. doi:10.1162/jocn_a_00002

Gathercole, S., Briscoe, J., Thorn, A., Tiffany, C., & ALSPAC, S. T. (2008). Deficits in verbal long-term memory and learning in children with poor phonological short-term memory skills. *The Quarterly Journal of Experimental Psychology*, *61*, 474-490. doi:10.1080/17470210701273443

Gathercole, S. E. (2006). Nonword Repetition and Word Learning: The Nature of the Relationship. *Applied Psycholinguistics*, *27*(04), 513-543. doi:10.1017/S0142716406060383

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, *28*(2), 200-213. doi:10.1016/0749-596X(89)90044-2

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology*, *33*(6), 966-979. doi:10.1037/0012-1649.33.6.966

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences*, *105*(37), 14222 -14227. doi:10.1073/pnas.0806530105

Gervain, J., & Mehler, J. (2010). Speech Perception and Language Acquisition in the First Year of Life. *Annual Review of Psychology*, *61*, 191-218. doi:10.1146/annurev.psych.093008.100408

Gervain, J., Mehler, J., Werker, J. F., Nelson, C. A., Csibra, G., Lloyd-Fox, S., Shukla, M., et al. (2011). Near-infrared spectroscopy: A report from the McDonnell infant methodology consortium. *Developmental Cognitive Neuroscience*, *1*(1), 22-46. doi:16/j.dcn.2010.07.004

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The Developmental Origins of Voice Processing in the Human Brain. *Neuron*, *65*(6), 852-858. doi:10.1016/j.neuron.2010.03.001

Gulya, M., Rovee-Collier, C., Galluccio, L., & Wilk, A. (1998). Memory Processing of a Serial List by Young Infants. *Psychological Science*, *9*, 303-307. doi:10.1111/1467-9280.00060

Gupta, P. (2005). Primacy and recency in nonword repetition. *Memory*, *13*, 318-324. doi:10.1080/09658210344000350

Gupta, P., Lipinski, J., Abbs, B., & Lin, P.-H. (2005). Serial position effects in nonword repetition. *Journal of Memory and Language*, *53*(1), 141-162. doi:10.1016/j.jml.2004.12.002

Habib, R., Nyberg, L., & Tulving, E. (2003). Hemispheric asymmetries of memory: the HERA model revisited. *Trends in Cognitive Sciences*, *7*(6), 241-245. doi:10.1016/S1364-6613(03)00110-4

Halle, M., & Vergnaud, J. (1990). *An essay on stress*. Cambridge, MA: MIT Press.

Havy, M., & Nazzi, T. (2009). Better Processing of Consonantal Over Vocalic Information in Word Learning at 16 Months of Age. *Infancy*, *14*(4), 439-456. doi:10.1080/15250000902996532

Hayes, B. (1995). *Metrical stress theory: principles and case studies*. University of Chicago Press.

Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(6), 644-655. doi:10.1016/S0022-5371(74)80052-6

Henson, R., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *The Quarterly Journal of Experimental Psychology Section A*, *56*, 1307-1334. doi:10.1080/02724980244000747

Henson, R. N. A. (1998). Short-Term Memory for Serial Order: The Start-End Model. *Cognitive Psychology*, *36*(2), 73-137. doi:10.1006/cogp.1998.0685

Henson, R. N. A. (1999). Positional information in short-term memory: Relative or absolute? *Memory & Cognition*, *27*, 915-927. doi:10.3758/BF03198544

Hochmann, J.-R. (2010, December 6). *Categories, words and rules in language acquisition* (Ph.D. thesis). Retrieved from http://digitallibrary.sissa.it/handle/1963/5005

Hochmann, J.-R., Benavides-Varela, S., Nespor, M., & Mehler, J. (2011). Consonants and vowels: different roles in early language acquisition. *Developmental Science*, *14*(6), 1445-1458. doi:10.1111/j.1467-7687.2011.01089.x

Homae, F., Watanabe, H., Nakano, T., Asakawa, K., & Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. *Neuroscience Research*, *54*(4), 276-280. doi:16/j.neures.2005.12.006

Hulme, C., Maughan, S., & Brown, G. D. . (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*(6), 685-701. doi:10.1016/0749-596X(91)90032-F

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1-2), 1-19. doi:10.1016/0010-0277(91)90045-6

Jusczyk, P W, Jusczyk, A. M., Kennedy, L. J., Schomberg, T., & Koenig, N. (1995). Young infants' retention of information about bisyllabic utterances. *Journal of Experimental Psychology. Human Perception and Performance*, *21*(4), 822-836.

Jusczyk, Peter W. (2000). *The discovery of spoken language*. MIT Press.

Jusczyk, Peter W., & Hohne, E. A. (1997). Infants' Memory for Spoken Words. *Science*, *277*(5334), 1984 -1986. doi:10.1126/science.277.5334.1984

Jusczyk, Peter W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*, 1465-1476. doi:10.3758/BF03213111

Kager, R. (1995). The metrical theory of word stress. *The handbook of phonological theory* (pp. 367-402). Blackwell.

Káldy, Z., & Leslie, A. M. (2005). A memory span of one? Object identification in 6.5-month-old infants. *Cognition*, *97*(2), 153-177. doi:10.1016/j.cognition.2004.09.009

Keidel, J. L., Jenison, R. L., Kluender, K. R., & Seidenberg, M. S. (2007). Does Grammar Constrain Statistical Learning? *Psychological Science*, *18*(10), 922 - 923. doi:10.1111/j.1467-9280.2007.02001.x

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*(1), 55-68.

Kotilahti, K., Nissil??, I., Huotilainen, M., M??kel??, R., Gavrielides, N., Noponen, T., Bj??rkman, P., et al. (2005). Bilateral hemodynamic responses to auditory stimulation in newborn infants. *NeuroReport*, *16*(12), 1373-1377. doi:10.1097/01.wnr.0000175247.35837.15

Kotilahti, K., Nissilä, I., Näsi, T., Lipiäinen, L., Noponen, T., Meriläinen, P., Huotilainen, M., et al. (2010). Hemodynamic responses to speech and music in newborn infants. *Human Brain Mapping*, *31*(4), 595-603. doi:10.1002/hbm.20890

Kovács, Á. M. (2008). *Learning two languages simultaneously: mechanisms recruited for dealing with a mixed linguistic input.* (PhD. dissertation). SISSA, Trieste.

Kovács, Á. M., & Mehler, J. (2009). Flexible Learning of Multiple Speech Structures in Bilingual Infants. *Science*, *325*(5940), 611 -612. doi:10.1126/science.1173947

Kuhl, P K, & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, *100*(4 Pt 1), 2425-2438.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*(4577), 1138-1141. doi:10.1126/science.7146899

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606 -608. doi:10.1126/science.1736364

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat Rev Neurosci*, *5*(11), 831-843. doi:10.1038/nrn1533

Kuhl, P. K., & Meltzoff, A. N. (1984). The Intermodal Representation of Speech in Infants. *Infant Behavior and Development*, *7*(3), 361-381. doi:10.1016/S0163-6383(84)80050-8

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13-F21. doi:10.1111/j.1467-7687.2006.00468.x

Kushnerenko, E., Cheour, M., Ceponiene, R., Fellman, V., Renlund, M., Soininen, K., Alku, P., et al. (2001). Central Auditory Processing of Durational Changes in Complex Speech Patterns by Newborns: An Event-Related Brain Potential Study. *Developmental Neuropsychology*, *19*(1), 83-97. doi:10.1207/S15326942DN1901_6

Lechner, H. A., Squire, L. R., & Byrne, J. H. (1999). 100 Years of Consolidation—Remembering Müller and Pilzecker. *Learning & Memory*, *6*(2), 77 -87. doi:10.1101/lm.6.2.77

Lee, H.-W., Rayner, K., & Pollatsek, A. (2001). The Relative Contribution of Consonants and Vowels to Word Identification during Reading. *Journal of Memory and Language*, *44*(2), 189-205. doi:06/jmla.2000.2725

Lee, H.-W., Rayner, K., & Pollatsek, A. (2002). The processing of consonants and vowels in reading: Evidence from the fast priming paradigm. *Psychonomic Bulletin & Review*, *9*(4), 766-772. doi:10.3758/BF03196333

Lenneberg, E. (1967). *Biological Foundations of Language*. New York: Wiley.

Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy. *Neuroscience and Biobehavioral Reviews*, *34*(3), 269-284.

Macchi Cassia, V., Simion, F., & Umiltaà, C. (2001). Face preference at birth: the role of an orienting mechanism. *Developmental Science*, *4*(1), 101-108. doi:10.1111/1467-7687.00154

Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*(6), 828-835. doi:10.1016/S0022-5371(69)80050-2

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' Cry Melody Is Shaped by Their Native Language. *Current Biology*, *19*(23), 1994-1997. doi:16/j.cub.2009.09.064

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' Recognition of the Sound Patterns of Their Own Names. *Psychological Science*, *6*(5), 314-317.

Marchetto, E. (2009). *Discovering words and rules from speech input: an investigation into early morphosyntactic acquisition mechanisms*. (Ph.D. Thesis). SISSA, Trieste.

Mattys, S. L., & Jusczyk, P. W. (2000). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91-121. doi:10.1016/S0010-0277(00)00109-8

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and Prosodic Effects on Word Segmentation in Infants. *Cognitive Psychology*, *38*(4), 465-494. doi:10.1006/cogp.1999.0721

McGaugh, J. L. (2000). Memory--a Century of Consolidation. *Science*, *287*(5451), 248 -251. doi:10.1126/science.287.5451.248

Meek, J. (2002). Basic principles of optical imaging and application to the study of infant development. *Developmental Science*, *5*(3), 371-380. doi:10.1111/1467-7687.00376

Mehler, J., Bertoncini, J., Barrière, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, *7*(5), 491 – 497. doi:10.1068/p070491

Mehler, J., Dehaene-Lambertz, G., Dupoux, E., & Nazzi, T. (1996). Coping with the Linguistic Diversity: The infants' viewpoint. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 101-116). New Jersey: Hillsdale.

Mehler, J., & Dupoux, E. (1994). *What Infants Know: The New Cognitive Science of Early Development.* Blackwell Publishers. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED379059

Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping*, *17*(2), 73-88. doi:10.1002/hbm.10042

Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language*, *89*(2), 277-289. doi:16/S0093-934X(03)00350-X

Miceli, G., Capasso, R., Benvegnu, B., & Caramazza, A. (2004). The Categorical Distinction of Vowel and Consonant Representations: Evidence from Dysgraphia. *Neurocase*, *10*(2), 109-121. doi:10.1080/13554790409609942

Miller, G. A., & Selfridge, J. A. (1950). Verbal Context and the Recall of Meaningful Material. *The American Journal of Psychology*, *63*(2), 176-185. doi:10.2307/1418920

Minagawa-Kawai, Y., Mori, K., Hebden, J. C., & Dupoux, E. (2008). Optical imaging of infants' neurocognitive development: Recent advances and perspectives. *Developmental Neurobiology*, *68*(6), 712-728. doi:10.1002/dneu.20618

Moon, C., Cooper, R. P., & Fifer, W. P. (October). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*(4), 495-500. doi:16/0163-6383(93)80007-U

Morgan, J. L., & Demuth, K. (1996). *Signal to syntax: bootstrapping from speech to grammar in early acquisition*. Routledge.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165-178.

Morton, J., & Chambers, S. M. (1976). Some Evidence for "speech" as an Acoustic Feature. *British Journal of Psychology*, *67*(1), 31-45. doi:10.1111/j.2044-8295.1976.tb01495.x

Morton, J., Hammersley, R. H., & Bekerian, D. A. (1985). Headed records: A model for memory and its failures. *Cognition*, *20*(1), 1-23. doi:10.1016/0010-0277(85)90002-2

Moscovitch, M. (Ed.). (1984). *Infant Memory, Its relation to normal and pathological memory in humans and other animals*. Advances in the Study of communication and affect (Vols. 1-10, Vol. 9). New York, NY, US: Plenum Press.

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, *78*, 679-684. doi:10.1037/h0026641

Nadel, L., & Zola-Morgan, S. (1984). Infantile amnesia: a neurobiological perspective. *Advances in the Study of communication and effect: Infant memory* (Vol. 9, pp. 145-172). New York, NY, US: Plenum Press.

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: differences between consonants and vowels. *Cognition*, *98*(1), 13-30. doi:16/j.cognition.2004.10.005

Nazzi, T., & Bertoncini, J. (2009). Phonetic Specificity in Early Lexical Acquisition: New Evidence from Consonants in Coda Positions. *Language and Speech*, *52*(4), 463 -480. doi:10.1177/0023830909336584

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756-766. doi:10.1037/0096-1523.24.3.756

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development*, *21*(4), 779-784. doi:16/S0163-6383(98)90044-3

Nazzi, T., Floccia, C., Moquet, B., & Butler, J. (2009). Bias for consonantal information over vocalic information in 30-month-olds: Cross-linguistic evidence from French and English. *Journal of Experimental Child Psychology*, *102*(4), 522-537. doi:16/j.jecp.2008.05.003

Nazzi, T., & New, B. (2007). Beyond stop consonants: Consonantal specificity in early lexical acquisition. *Cognitive Development*, *22*(2), 271-279. doi:16/j.cogdev.2006.10.007

Nelson, C. A. (1995). The ontogeny of human memory: A cognitive neuroscience perspective. *Developmental Psychology*, *31*(5), 723-738. doi:10.1037/0012-1649.31.5.723

Nespor, M., Peña, M., & Mehler, J. (2003). On the Different Roles of Vowels and Consonants in Speech Processing and Language Acquisition. *Lingue e linguaggio*, (2), 203–230.

Nespor, M, Guasti, M. T., & Christophe, A. (1996). Selecting word order: The rythmic activation principle. *Interfaces in phonology*. Akademie Verlag.

Nespor, M, & Vogel, I. (2007). *Prosodic phonology*. Walter de Gruyter.

New, B., Araújo, V., & Nazzi, T. (2008). Differential Processing of Consonants and Vowels in Lexical Access Through Reading. *Psychological Science*, *19*(12), 1223 -1227. doi:10.1111/j.1467-9280.2008.02228.x

Newport, E. L. (1990). Maturational Constraints on Language Learning. *Cognitive Science*, *14*(1), 11-28. doi:10.1207/s15516709cog1401_2

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*(1), 1-25. doi:10.1002/hbm.1058

Nyberg, L., Cabeza, R., & Tulving, E. (1996). PET studies of encoding and retrieval: The HERA model. *Psychonomic Bulletin & Review*, *3*, 135-148. doi:10.3758/BF03212412

Oakes, L. M., & Bauer, P. J. (2007). *Short-and long-term memory in infancy and early childhood: taking the first steps toward remembering*. Oxford University Press.

Obrig, H., Rossi, S., Telkemeyer, S., & Wartenburger, I. (2010). From acoustic segmentation to language processing: evidence from optical imaging. *Frontiers in Neuroenergetics*, *2*, 13. doi:10.3389/fnene.2010.00013

Olson, I. R., & Berryhill, M. (2009). Some surprising findings on the involvement of the parietal lobe in human memory. *Neurobiology of Learning and Memory*, *91*(2), 155-165. doi:10.1016/j.nlm.2008.09.006

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*(4), 761-781. doi:10.1037/0033-295X.105.4.761-781

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, *30*(3), 331-347. doi:10.1016/0749-596X(91)90040-Q

Papagno, C., & Vallar, G. (1992). Phonological Short-term Memory and the Learning of Novel Words: The Effect of Phonological Similarity and Item Length. *The Quarterly Journal of Experimental Psychology Section A*, *44*, 47-67. doi:10.1080/14640749208401283

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674-681. doi:10.1038/nn1082

Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford University Press.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-Driven Computations in Speech Processing. *Science*, *298*(5593), 604 -607. doi:10.1126/science.1072901

Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences*, *100*(20), 11702 -11705. doi:10.1073/pnas.1934290100

Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., Poloniato, A., et al. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences*, *108*(38), 16056 -16061. doi:10.1073/pnas.1102991108

Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., Baldoli, C., et al. (2010). Functional specializations for music processing in the human newborn brain. *Proceedings of the National Academy of Sciences*, *107*(10), 4758 -4763. doi:10.1073/pnas.0909074107

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition*, *100*(1), 1-32. doi:10.1016/j.cognition.2005.11.004

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nat Neurosci*, *6*(7), 688-691. doi:10.1038/nn1083

Piaget, J. (1952). *The origins of intelligence in children.* (M. Cook, Trans.). New York, NY, US: W W Norton & Co. Retrieved from http://psycnet.apa.org/psycinfo/2007-10742-000/

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*, 253-260. doi:10.3758/BF03214136

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421-435. doi:10.1037/0096-1523.20.2.421

Pons, F., & Toro, J. M. (2010). Structural generalizations over consonants and vowels in 11-month-old infants. *Cognition*, *116*(3), 361-367. doi:10.1016/j.cognition.2010.05.013

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Vervoort, P. (1988). Intra-amniotic transmission of the human voice. *Revue Française De Gynécologie Et D'obstétrique*, *83*(1), 43-50.

Richmond, J., & Nelson, C. A. (2007). Accounting for change in declarative memory: A cognitive neuroscience perspective. *Developmental Review*, *27*(3), 349-373. doi:10.1016/j.dr.2007.04.002

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339-349. doi:10.1111/j.1467-7687.2008.00786.x

Rost, G. C., & McMurray, B. (2010). Finding the Signal by Adding Noise: The Role of Noncontrastive Phonetic Variability in Early Word Learning. *Infancy*, *15*(6), 608-635. doi:10.1111/j.1532-7078.2010.00033.x

Rugg, M. D., Fletcher, P. C., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1996). Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain*, *119*(6), 2073 -2083. doi:10.1093/brain/119.6.2073

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926 -1928. doi:10.1126/science.274.5294.1926

Schacter, D. L. (2002). *The Seven Sins of Memory: How the Mind Forgets and Remembers*. Houghton Mifflin Harcourt.

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: edge alignment facilitates target extraction. *Developmental Science*, *9*(6), 565-573. doi:10.1111/j.1467-7687.2006.00534.x

Shadmehr, R., & Holcomb, H. H. (1997). Neural Correlates of Motor Memory Consolidation. *Science*, *277*(5327), 821 -825. doi:10.1126/science.277.5327.821

Shallice, T., Fletcher, P., Frith, C. D., Grasby, P., Frackowiak, R. S. J., & Dolan, R. J. (1994). Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature*, *368*(6472), 633-635. doi:10.1038/368633a0

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology*, *22*, 261-273. doi:10.1080/00335557043000203

Shillinger, G. L., Palacios, D. M., Bailey, H., Bograd, S. J., Swithenbank, A. M., Gaspar, P., Wallace, B. P., et al. (2008). Persistent Leatherback Turtle Migrations Present Opportunities for Conservation. *PLoS Biol*, *6*(7), e171. doi:10.1371/journal.pbio.0060171

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*(1), 1-32. doi:10.1016/j.cogpsych.2006.04.002

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, *108*(15), 6038 -6043. doi:10.1073/pnas.1017617108

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, *105*(2), 809 -813. doi:10.1073/pnas.0707021105

Sinclair, H. (1973). Language acquisition and cognitive development. In T. E. Moore (Ed.), *Cognitive Development and the Acquisition of Language.* New York, NY, US: Academic Press.

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, *106*(2), 833-870. doi:10.1016/j.cognition.2007.05.002

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*(2), 195-231. doi:10.1037/0033-295X.99.2.195

Stefanics, G., Háden, G. P., Sziller, I., Balázs, L., Beke, A., & Winkler, I. (2009). Newborn infants process pitch intervals. *Clinical Neurophysiology*, *120*(2), 304-308. doi:16/j.clinph.2008.11.020

Swain, I. U., Zelazo, P. R., & Clifton, R. K. (1993). Newborn infants' memory for speech sounds retained over 24 hours. *Developmental Psychology*, *29*(2), 312-323. doi:10.1037/0012-1649.29.2.312

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3617 -3632. doi:10.1098/rstb.2009.0107

Taga, G., & Asakawa, K. (2007). Selectivity and localization of cortical response to auditory and visual stimulation in awake infants aged 2 to 4 months. *NeuroImage*, *36*(4), 1246-1252. doi:10.1016/j.neuroimage.2007.04.037

Taylor, H. C. (1933). The fundamental pitch of English vowels. *Journal of Experimental Psychology*, *16*(4), 565-582. doi:37/h0070672

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., Obrig, H., et al. (2009). Sensitivity of Newborn Auditory Cortex to the Temporal Structure of Sounds. *The Journal of Neuroscience*, *29*(47), 14726 - 14733. doi:10.1523/JNEUROSCI.1246-09.2009

Thatcher, R., Walker, R., & Giudice, S. (1987). Human cerebral hemispheres develop at different rates and ages. *Science*, *236*(4805), 1110 -1113. doi:10.1126/science.3576224

Tincoff, R., & Jusczyk, P. W. (1999). Some Beginnings of Word Comprehension in 6-Month-Olds. *Psychological Science*, *10*(2), 172 -175. doi:10.1111/1467-9280.00127

Toro, J. M., Shukla, M., Nespor, M., & Endress, A. D. (2008). The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences. *Perception & Psychophysics*, *70*(8), 1515-1525. doi:10.3758/PP.70.8.1515

Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding Words and Rules in a Speech Stream. *Psychological Science*, *19*(2), 137 -144. doi:10.1111/j.1467-9280.2008.02059.x

Trehub, S. E. (2001). Musical Predispositions in Infancy. *Annals of the New York Academy of Sciences*, *930*(1), 1-16. doi:10.1111/j.1749-6632.2001.tb05721.x

Trehub, S. E. (2003). The developmental origins of musicality. *Nat Neurosci*, *6*(7), 669-673. doi:10.1038/nn1084

Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, *92*(1-2), 231-270. doi:10.1016/j.cognition.2003.10.008

Valenza, E., Simion, F., Cassia, V. M., & Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 892-903. doi:10.1037/0096-1523.22.4.892

Valiante, A. G., Barr, R. G., Zelazo, P. R., Papageorgiou, A. N., & Young, S. N. (2006). A Typical Feeding Enhances Memory for Spoken Words in Healthy 2- to 3-Day-Old Newborns. *Pediatrics*, *117*(3), e476 -e486. doi:10.1542/peds.2004-2859

Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, *24*(5), 573-583. doi:10.3758/BF03201084

Villringer, A., & Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences*, *20*(10), 435-442. doi:10.1016/S0166-2236(97)01132-6

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science*, *7*(3), 270-276.

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*(9), 445-453. doi:10.1016/j.tics.2005.07.001

Werker, J F, & Tees, R. C. (1999). Influences on infant speech processing: toward a new synthesis. *Annual Review of Psychology*, *50*, 509-535. doi:10.1146/annurev.psych.50.1.509

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49-63. doi:16/S0163-6383(84)80022-3

Zatorre, R. J., & Belin, P. (2001). Spectral and Temporal Processing in Human Auditory Cortex. *Cerebral Cortex*, *11*(10), 946 -953. doi:10.1093/cercor/11.10.946

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, *6*(1), 37-46. doi:10.1016/S1364-6613(00)01816-7

Appendices

During my experience as PhD. student, the Language Cognition and Development laboratory offered me the opportunity to develop and get involved in various studies. Some of these studies inspired the thesis that was presented above, and some others were only partially related to the main thesis project. The additional studies constitute the basis for these appendices. Many of these studies were the result of fruitful discussions and collaborations with different colleagues about topics that had and continue to be of the interest of the laboratory. These studies served as pilots, as preliminary data to sustain methodological discussions, or as appetizers to begin testing various topics related to cognitive development such as auditory perception, language universals, bilingualism, etc.

From these discussions in the laboratory we have developed some ideas and studies that have been recently published. These publications are presented in the first two appendices (A and B) of this thesis. In appendices C to F, I show some collected data that have not been published (maybe they won't ever be) that perhaps could be useful in discussions that may arise in future meetings in our own lab. The purpose of these appendixes is precisely to provide the laboratory, and the new members, with a record of the information that we have obtained in pilot studies, hoping that it could serve as an internal reminder, and that it could save some time to future explorations that students and colleagues in the lab intend to carry out in the same topics.

Appendix A. The Iambic-trochaic law and its effects on early word segmentation

Article

Language and Speech

Language *and* Speech
54(1) 123–140

DOI: 10.1177/0023830910388018
las.sagepub.com
SAGE

Acoustic Markers of Prominence Influence Infants' and Adults' Segmentation of Speech Sequences

Ricardo A. H. Bion[1,2], Silvia Benavides-Varela[2], Marina Nespor[2,3,4]

Abstract

Two experiments investigated the way acoustic markers of prominence influence the grouping of speech sequences by adults and 7-month-old infants. In the first experiment, adults were familiarized with and asked to memorize sequences of adjacent syllables that alternated in either pitch or duration. During the test phase, participants heard pairs of syllables with constant pitch and duration and were asked whether the syllables had appeared adjacently during familiarization. Adults were better at remembering pairs of syllables that during familiarization had short syllables preceding long syllables, or high-pitched syllables preceding low-pitched syllables. In the second experiment, infants were familiarized and tested with similar stimuli as in the first experiment, and their preference for pairs of syllables was accessed using the head-turn preference paradigm. When familiarized with syllables alternating in pitch, infants showed a preference to listen to pairs of syllables that had high pitch in the first syllable. However, no preference was found when the familiarization stream alternated in duration. It is proposed that these perceptual biases help infants and adults find linguistic units in the continuous speech stream. While the bias for grouping based on pitch appears early in development, biases for durational grouping might rely on more extensive linguistic experience.

Keywords

language development, memory, perceptual biases, speech segmentation

Introduction

When learning a language, infants are faced with the task of finding words and phrases, and of discovering the relation between these units in an apparently continuous stream. In order to get the

---

[1] Stanford University
[2] International School for Advanced Studies ISAS-SISSA, Trieste
[3] University of Milan-Bicocca
[4] Accademia dei Lincei

Corresponding author:
Ricardo A. H. Bion, Department of Psychology, Stanford University, Jordan Hall, Building 420, Stanford, CA 94305, USA
Email: ricardoh@stanford.edu

analysis started, predispositions for paying attention to specific acoustic cues, tendencies to generalize, and constraints against overgeneralization must be present at a very early age. Discovering the way these abilities develop from infancy to adulthood is one of the main challenges for language acquisition research.

Several studies indicate that infants' first linguistic preferences and discrimination abilities are guided by the prosodic and rhythmic properties of speech. Some days after birth, infants rely on these properties to distinguish between different languages (Nazzi, Bertoncini, and Mehler, 1998; Ramus, Hauser, Miller, Morris, and Mehler, 2000). They also appear to exploit them to identify their native language and their mother's voice (De Casper and Fifer, 1980). Within their first months of life, infants make use of prosody to find words, clauses, and phrases in fluent speech (Christophe, Gout, Peperkamp, and Morgan, 2003; Gout, Christophe, and Morgan, 2004; Jusczyk, Cutler, and Redanz, 1993; Jusczyk et al., 1992). The correlations between linguistic units and prosodic cues (e.g., Cooper and Paccia-Cooper, 1980; Goldman-Eisler, 1972), and infants' early sensitivity to the latter, have led to the proposal that acoustic information, such as variations in duration and pitch, might help infants break into language (for a review, Morgan and Demuth, 1996; Weissenborn and Höhle, 2001a, 2001b).

A central issue in the field of language acquisition concerns the extent to which these perceptual biases are innate or emerge from linguistic experience. One of the earliest prosodic biases observed is a preference for trochaic over iambic words in early infancy. English-learning 9-month-olds prefer to listen to lists of trochees over lists of iambs (Jusczyk et al., 1993), and German-learning infants show the same preference as early as 6 months of age (Höhle, Bijeljac-Babicb, Weissenborn, and Nazzi, 2009). At a similar point in development, Dutch, German, and English learning infants have a trochaic bias in the segmentation of words from fluent speech (Cutler, Mehler, Norris, and Segui, 1986; Cutler and Norris, 1988; Höhle, 2002; Houston, Jusczyk, Kuijpers, Coolen, and Cutler, 2000).

Three different explanations for infants' early trochaic biases are possible. One of the possibilities is that the trochaic bias is innate and universal (Allen and Hawkins, 1978), either being present at birth, or emerging between 6 and 9 months of age, when it is observed in different languages (Höhle et al., 2009; Jusczyk et al., 1993). According to this proposal, mistakes in segmentation will occur in case infants are learning a language that does not follow a trochaic pattern at the word level. A second possibility is that the trochaic bias emerges after generalizations are extracted over isolated words or over words segmented using distributional analyses (Thiessen and Saffran, 2003). This proposal is supported by experiments that show that when prosodic and distributional cues are contrasted, 7-month-old infants prefer statistically coherent words at the segmental level over trochaic words (Thiessen and Saffran, 2003). A third possibility is that trochaic biases only emerge for infants learning stress-based languages such as German, Dutch, or English, and should thus not be present in cases where infants are learning syllable-timed languages such as French or Italian (Nazzi, Iakimova, Bertoncini, Frédonie, and Alcantara, 2006).

This third proposal, known as the rhythmic-activation proposal, is supported by two different pieces of evidence. The first supporting evidence is that infants are sensitive to the rhythmic properties of languages from birth onwards (Nazzi, Bertoncini, and Mehler, 1998; Ramus et al., 2000), and therefore have the perceptual mechanisms necessary to trigger the trochaic bias. The second source of evidence comes from cross-linguistic studies that show that infants learning French, a syllabic language, segment words using a syllabic strategy, while infants learning English, a stress-based language, rely on a trochaic bias instead (Nazzi et al., 2006). Additional evidence comes from the fact that infants learning German prefer trochaic over iambic words at 6 months of age, while this bias is not present in French-learning infants of the same age (Höhle et al., 2009). In addition, studies using event-related potentials find different neural responses to trochaic and iambic words in French- and German-learning infants as early as 4 months of age (Friederici, Friedrich, and Christophe, 2007).

However, these studies fail to acknowledge that stress is marked by different acoustic cues in different languages, and that infants' grouping biases might rely on acoustic cues rather than on adults' perception of the position of lexical stress. Most of the acoustic and perceptual descriptions of word stress focused on Germanic stress-timed languages, and indicated that word stress is marked mainly by changes in pitch and duration, and less reliably by changes in intensity (Cutler, 2005). It is generally agreed that in English, none of these acoustic cues separately is crucial, but that all are evaluated simultaneously when judging lexical stress (Lieberman, 1960). However, in other languages, these cues have been shown to be orthogonal. In Welsh, strong effects of duration, but inconsistent effects of pitch, are found in the perception of lexical stress (Williams, 1985). In Thai and other tone languages, stress is signaled uniquely by duration (Potisuk, Gandour, and Harper, 1996), while in Polish stress is signaled uniquely by pitch (Dogil, 1999). Even when the same acoustic cues are used, their exact weighting might change across languages (Bleakley, 1973; Llisterri, Machuca, de la Mota, Riera, and Rios, 2003) and even across speakers of the same language (Nakatani and Aston, 1978). Therefore, talking about a trochaic universal bias without taking into account the way stress is realized at the acoustic levels across languages is problematic. This fact is aggravated by the fact that in many cross-linguistic studies, stimuli from one single language are used to test perceptual biases in several different languages (Friederici et al., 2007; Höhle et al., 2009).

Different acoustic realizations of stress are also present at other linguistic levels. French, for example, does not have consistent stress at the lexical level, but is an iambic language at the level of the phonological phrase, with stress marked mainly by increased duration (Dupoux, Pallier, Sebastián-Gallés, and Mehler, 1997). Even within languages of the same rhythmic class, different acoustic cues are used to signal clause boundaries. In English, for example, stress is marked by higher resets in pitch than in Dutch, which relies more heavily on pauses (Johnson and Seidl, 2008). Interestingly, Johnson and Seidl (2008) found that Dutch- and English-learning infants rely on different acoustic cues to segment clauses (Johnson and Seidl, 2008). However, in their conclusions, the authors point out that two explanations are possible for this pattern of results: on the one hand, it is possible that the infants weight acoustic cues differently; on the other hand, it is possible that infants use a universally similar cue weighting, but that clause boundaries are not as saliently marked by pitch and duration in Dutch. A similar acknowledgement should be made by studies discussing infants' early trochaic biases, with special emphasis on cross-linguistic variation and on the influence of acoustic cues on different levels of linguistic processing, such as the word and the phonological phrase level, on infants' early biases.

Referring back to perceptual biases for grouping, studies with adults show that iambic and trochaic biases can be switched depending on the acoustic cue being manipulated. In short, humans segment sequences of sounds that differ only in duration with the longest element in final position (iambs) and sequences of sounds that differ only in intensity with the more intense element in initial position (trochees). This bias was initially proposed to explain the grouping of musical sequences (Bolton, 1894; Cooper and Meyer, 1960; Woodrow, 1951), but was later expanded to account for regularities in speech production and biases in speech perception.

The first formulations of the iambic-trochaic law for language focused on the word level and on the effects of duration and intensity on the grouping of syllables. In speech production, this law was proposed to account for the location of word secondary stress crosslinguistically. Specifically, if secondary stresses are initial in the phonological constituent known as the metrical foot, then they are manifested mainly with increased intensity; if they are final, they are manifested mainly with increased duration (Hayes, 1995). The existence of the iambic-trochaic grouping principle is confirmed by experimental evidence. Hay and Diehl (2007), for example, investigated grouping preferences for linguistic and non-linguistic stimuli by French and English speakers. Listeners were presented a sequence of repeating syllables or tones varying in either intensity or duration, and had to decide whether the sequence was formed by trochaic or iambic sub-sequences. Results from both modalities (speech and non-speech) and from both languages (English and French) corroborated the iambic-trochaic law, supporting previous claims that it must be guided by general principles of the human auditory system (Bolton, 1894; Cooper and Meyer, 1960; Woodrow, 1951).

The most recent extensions of the iambic-trochaic law focused on the phrasal level and on the role of duration and pitch. In speech production, it has been proposed that the location of stress in phrasal constituents differs depending on whether it is initial or final, i.e. on the first or on the last word of the phonological phrase (Nespor and Vogel, 2008). It has then been proposed that the physical realization of phonological phrase stress also varies depending on its location: increased duration characterizes iambic patterns, while greater intensity and higher pitch characterize trochaic patterns (Nespor, Shukla, van de Vijver, Avesani, Schraudolf, and Donati, 2008). These findings, however, should be replicated with infant-directed speech. Still, from this body of studies, it is possible to conclude that in speech production, units with higher pitch or intensity mark the beginning of speech sequences, whereas units having long duration mark the end of speech sequences. However, it was not investigated whether pitch would also lead to trochaic grouping in speech perception as previously shown for intensity (Hay and Diehl, 2007).

Interestingly, Nespor et al. (2008) propose that the different manifestation of prominence within phonological phrases might be exploited to acquire word order (cf. also Nespor, Guasti, and Christophe, 1996). Languages with initial prominence (trochaic prominence marked by pitch) within phonological phrases have syntactic heads that follow their complements, and languages with final prominence (iambic prominence marked by duration) within phonological phrases have syntactic heads that precede their complements (Nespor and Vogel, 2008). Broadly, head-direction refers to the ordering of complements with respect to their head: for example, objects with respect to either verbs or adpositions, or nominal complements with respect to nouns. In head-initial languages, like English and French, for example, verbs precede objects, while in head-final languages, like Turkish and Japanese, verbs follow objects. In addition, in head-initial languages, function words tend to precede content words, while in head-final languages function words tend to follow them. Importantly, Nespor et al. (2008) also found that in German, a language in which the object can either follow or precede the verb or the adposition depending on syntactic factors, the prominence of the phonological phrase is differently realized: mainly through duration, as in French, when the object is final and mainly through pitch, as in Turkish, when the object is initial. The iambic-trochaic law could thus also be of aid to infants exposed to a language with mixed word order.

This theoretical proposal became more appealing when Christophe, Guasti, Nespor, and van Ooyen (2003) showed that infants are sensitive to these prosodic differences. Infants of 6 to 12 weeks of age can discriminate between phrases with different orders of heads and complements; specifically, they can distinguish French (head-initial) from Turkish (head-final) phrases. Importantly, the sentences in this study were resynthesized to remove phonemic information, and were also matched on all factors that are thought to influence prosody, differing only in their stress pattern within phonological phrases. Namely, French follows the prominence final pattern of head-initial languages, and Turkish follows the prominence initial pattern of head-final languages.

In addition to evidence across languages, infants were also shown to be sensitive to prosodic cues that signal the order of heads and complements within a single language. Bion, Höhle, and Schmitz (2007) found that 14-month-old German infants distinguish between head-initial and head-final German phrases, but that this ability is no longer present when the prosodic information is not isomorphic with the ordering of words within phrases. In these studies, however, the exact acoustic cues on which infants were relying for discriminating phrases within the same language (Bion et al., 2007) and across languages (Christophe et al., 2003) were not manipulated, and infants' preferences could be guided by variations in intensity, pitch, or duration. In music perception, 4.5-month-old infants were shown to rely on longer duration and lower pitch to perceive boundaries of tonal groups (Jusczyk and Krumhansl, 1993). However, once more, pitch and duration were present simultaneously in the stream, and it is not possible to determine their use in isolation.

In the present study, we aimed at investigating this issue further, focusing on how acoustic markers of prominence influence the grouping of linguistic units by infants and adults. The current study is different from the previous studies discussed so far in several respects. First, the previous studies with adults used a direct measurement of perceptual grouping, presenting a sequence of syllables or tones to the listeners, and then asking whether participants had heard iambic or trochaic sequences (Hay and Diehl, 2007). Since subjects heard the same syllable or tone repeated several times, having to decide on an internal ordering might be considered an artificial task. In the present study, we aim to investigate perceptual grouping indirectly, by focusing on whether linguistic units congruent with the iambic-trochaic law are better remembered than incongruent units, in the case of adults, or preferred over incongruent units, in the case of infants. Studies on event segmentation in the visual domain show that event segmentation is automatic and is best observed by implicit behavioral and neurophysiological measures. Often, direct measures, such as asking subjects about event boundaries, can influence the perceptual processes they attempt to measure (Zacks and Swallow, 2007). Studies on visual segmentation indicate that memory for events is a robust implicit measure of event segmentation (Kurby and Zacks, 2008; Swallow and Zacks, 2008; Swallow, Zacks, and Abrams, 2009). An additional advantage of using an indirect measure with adults lies in the possibility of better comparing the results from the adult and infant experiments. In the case of the infant experiment, it is, in fact, obviously not possible to ask participants for conscious judgments of event boundaries.

Another fundamental difference between our study and previous ones lies in the acoustic markers of prominence under investigation. Specifically, while previous studies focused on the effect of intensity in rhythmic grouping, the present study focuses on pitch instead. It is hard to dissociate the effect of pitch from the effect of intensity, since sounds with higher pitch tend to be perceived as having higher intensity. However, it has been shown that, at the onset of lexical recognition, infants are sensitive to variations in pitch but ignore variations in intensity (Singh, White, and Morgan, 2008). This strategy is particularly profitable, as amplitude variations are highly influenced by the listening environment, such as the distance or the presence of a barrier between speaker and listener. In contrast, pitch is modulated solely by the speaker. While most languages use pitch to encode linguistic information, no language is known to rely only on intensity to signal linguistic contrasts. These pieces of evidence taken together appear to indicate that intensity might be a very poor candidate to guide infants' early perceptual biases.

The biggest difference between our and previous studies, however, lies in the populations under investigation. To our knowledge, this is the first study to use a similar set of familiarization and test stimuli to test both infants and adults. The developmental nature of our study can bring insights into the amount of linguistic exposure necessary for the emergence of the iambic-trochaic biases. Specifically, two studies were carried out: the first investigates the role of alternations in pitch or duration on memory of speech sequences by Italian adults, while the second investigates the effect of these same markers of prominence on the listening preferences of Italian 7-month-old infants.

We predict that adult native speakers will group the sequences as duration-final and pitch-initial. It is an open question how perceptual grouping will work with infants. The rhythmic-activation proposal (Nazzi et al., 2006, not to be confused with Nespor and colleagues' rhythmic-activation principle, 1996) would predict neither an iambic nor a trochaic bias in our infants, since Italian is a syllable-based language and the trochaic bias should be present only in infants learning stress-timed languages. In contrast, proposals that claim that the trochaic bias is universal, and do not take into consideration the role of different acoustic cues (Allen and Hawkins, 1978), would predict trochaic biases both for streams varying in pitch and for streams varying in duration. Alternatively, it is possible that the iambic-trochaic bias is present early in development. In this case, infants should segment streams alternating in pitch as formed by trochees, and streams alternating in duration as formed by iambs. Last, it is possible that perceptual biases relying on pitch and on duration emerge at different points in development.

2 Experiment 1

2.1 Subjects

Participants were 45 Italian adults, 15 in each experimental condition (duration: 7 males, mean age = 23.8, range = 21–29; pitch: 5 males, mean age = 22.4, range = 19–26; control: 8 males, mean age = 24.1, range = 22–27).

2.2 Stimuli

This experiment consisted of three different conditions. In the first condition (pitch condition), the syllables in the familiarization stream alternated in pitch; in the second condition (duration condition), syllables alternated in duration; and in the third condition (control condition), all syllables had the same pitch and duration (Figure 1).

Specifically, in the pitch condition, a stream with 10 adjacent syllables (/pa su tu ke ma vi bu go ne du/), separated by an interval of 100 ms, was concatenated and repeated 20 times. The first two syllables had their pitch set to 180 Hz and their phoneme duration set to 180 ms. Starting from the third syllable, the pitch of the central portion of the vowel of the odd syllables was increased in 20 Hz steps, until it reached and remained at 400 Hz. Therefore, the third syllable had a pitch of 200 Hz (180 + 20), the fourth 180 Hz, the fifth 220 Hz (200 + 20), the sixth 180 Hz, until the odd syllable peaked and remained at 400 Hz. This gradual increase in pitch was implemented to avoid participants forming bisyllabic groupings from the onset of the familiarization, as previous studies showed that rhythmic judgments might be based on the pattern of the first two stimuli of a sequence (Hay and Diehl, 2007).

In the duration condition, the familiarization stream was composed of the same stream of syllables and acoustic cues as in the pitch condition. In this condition, however, instead of an increase in pitch, the odd syllables had prominence marked by increasing the duration of their vowel, in 20 ms steps, up to 400 ms. In the control condition, the same syllables were used, but no marker of prominence was added. The values for pitch and duration were extrapolated from recordings made at SISSA Language and Development lab. These values were based on acoustic measurements of productions of child-directed iambs and trochees, are within the range of values used in previous experiments (Hay and Diehl, 2007), and fall within the range of values naturally found in speech (Crystal and House, 1990). The three familiarization streams were synthesized with the female voice of the MBROLA Italian database IT4 (Dutoit, Pagel, Pierret, Bataille, and Vreken, 1996).

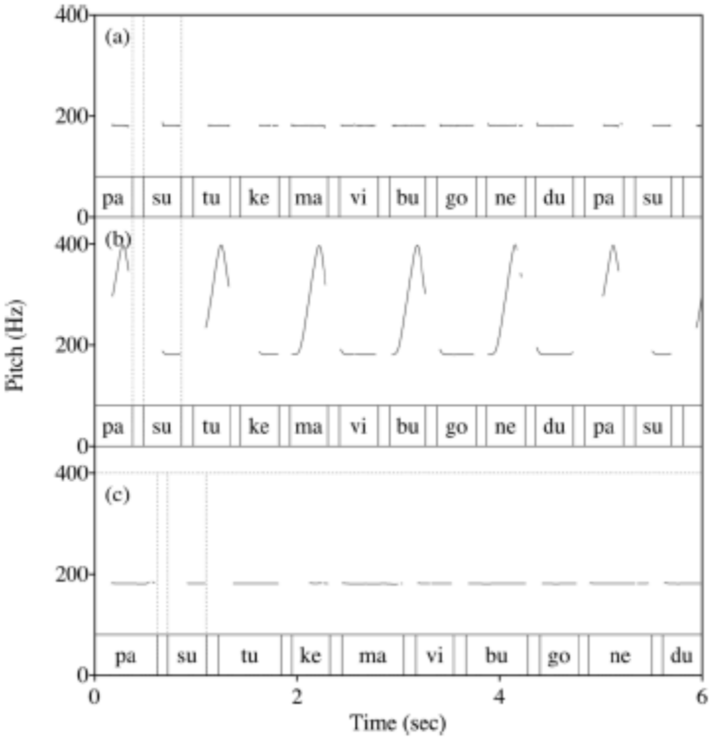

Figure 1. A segment from the middle of the familiarization stream with the three conditions of Experiment 1, (a) control condition, with phoneme duration of 180 ms and pitch of 180 Hz, (b) pitch condition, same as flat condition but with the pitch of the odd syllables peaking at 400 Hz at the midpoint of its vowel, (c) duration condition, same as control condition, but with the vowel of the odd syllables lasting 400 ms

For the test, three different sets of stimuli were synthesized by pairing syllables present in the familiarization. The first set of stimuli was created by pairing syllables that occurred adjacently and had prominence in their first syllable during familiarization (i.e., pa-su, tu-ke, ma-vi, bu-go, ne-du); the second set of stimuli was created by pairing syllables that occurred adjacently and had prominence on their second syllable during familiarization (i.e., su-tu, ke-ma, vi-bu, go-ne, du-pa); and the third by pairing non-adjacent syllables (i.e., pa-vi, vi-pa, su-bu, bu-su, tu-go, go-tu, ke-ne, ne-ke, ma-du, du-ma). The pairs of syllables that were synthesized were all spondees with pitch set to 180 Hz and phoneme duration set to 180 ms, and were the same for all participants. In this way, differences in performance could be attributed only to the differences in the familiarization streams.

2.3 Procedure

Participants were tested in a silent room and listened to the stimuli through headphones. After listening to one of the three possible familiarization streams, they were presented with the test stimuli, and were asked to judge whether they had heard these pairs of syllables adjacently during the familiarization phase or not. Each pair of syllables was presented twice, and a different random

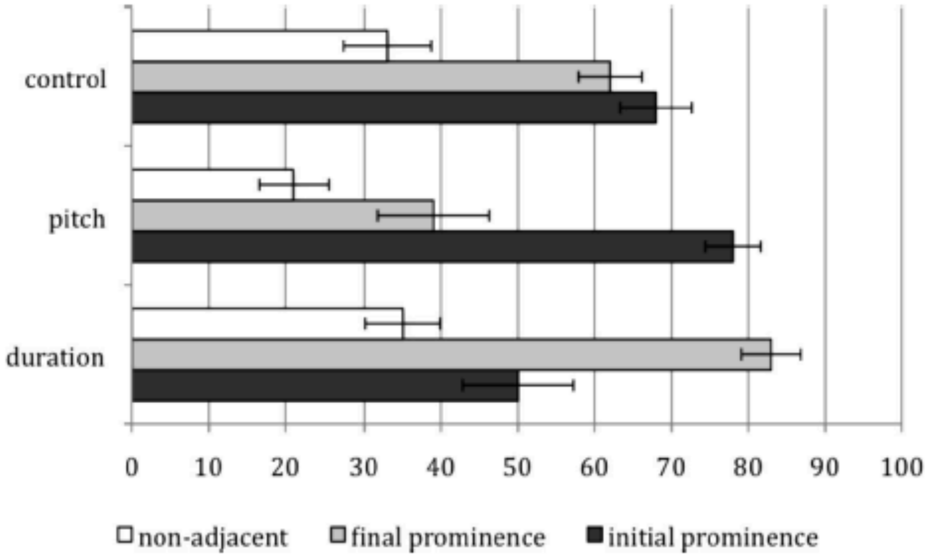

Figure 2. Percentage correct and standard error for the three experimental groups on the three sets of stimuli

order was created for each participant. The experiment was controlled by PsyScope X (Cohen, MacWhinney, Flatt, and Provost, 1993).

2.4 Results

In this experiment, participants in the pitch and in the duration conditions remembered different sets of stimuli during the test phase. An ANOVA indicates an interaction between familiarization (pitch, duration, and control) and condition (iamb, trochee, non-word) to be statistically significant, $F(1, 28) = 31.74$, $p < 0.001$, with no main effect of group, $F(1, 28) = 2.246$, $p = 0.145$, or test, $F(1, 28) = 0.225$, $p = 0.639$ (Figure 2). Specifically, participants in the pitch group were better at remembering pairs of syllables that had prominence in the first syllable during the familiarization than pairs of syllables that had prominence in their second syllables, $t(14) = 4.64$, $p < 0.001$. In contrast, participants in the duration group better remembered pairs of syllables with prominence in the second syllable, $t(14) = -3.426$, $p = 0.004$.

In order to confirm that these results were not caused by a particular property of the syllables chosen, the performance of the participants in the control group was analyzed. In this control condition, in which no acoustic marker of prominence was present during familiarization, no preference was observed for either set of stimuli, $t(14) = 0.84$, $p = 0.427$. In all three experimental groups, participants recalled pairs of syllables that occurred adjacently in familiarization significantly more often than pairs of non-adjacent syllables, $p < 0.05$.

These results strongly support our prediction that pairs of syllables alternating in pitch will be grouped with initial prominence (trochees), while sequences alternating in duration will be grouped with final prominence (iambs). Markers of prominence indicate the chunks over which the stream should be processed. In this experiment, instead of trying to memorize ten different syllables, participants were guided to memorize five adjacent pairs of syllables that followed the principles of the iambic-trochaic law. This pre-processing of the stimuli might diminish cognitive load and be highly profitable in speech processing and language acquisition (Miller, 1956).

In order to investigate whether the same pattern of perceptual grouping would also be present earlier in development, Experiment 2 focused on 7-month-old infants' listening preferences.

3 Experiment 2

3.1 Subjects

Participants were 48 Italian infants, 16 in each experimental condition (duration condition: age range 7m 05d–8m 02d, mean age 7m 17d; pitch condition: age range 7m 01d–8m 03d, mean age 7m 16d; control condition: age range 7m 01d–8m 05d, mean age 7m 17d). Seventeen additional infants were tested but not included in the analyses (duration condition: 7 infants rejected, 5 fussy, and 2 for accumulated rejected trials;[1] pitch condition: 6 infants rejected, 4 fussy, 2 for accumulated rejected trials; control condition: 4 infants rejected, 3 fussy, 1 experimental error).

3.2 Stimuli

As in Experiment 1, three different conditions were investigated. Again, the difference between the conditions lay in the acoustic cue used to signal prominence during the familiarization phase. In the pitch condition, syllables alternated in pitch; in the duration condition, they alternated in duration; and in the control condition, all syllables had the same pitch and duration.

The same acoustic markers of prominence and stimuli from Experiment 1 were used in this experiment. However, instead of 10 adjacent syllables, the familiarization consisted of 6 adjacent syllables (/pa su tu ke ma vi/). The familiarization stream was three minutes long.

For the test, six different pairs of adjacent syllables were synthesized, comprising two different sets of stimuli. The first set of stimuli was created by pairing syllables that had prominence in their first syllable during the familiarization (i.e., pa-su, tu-ke, ma-vi), while the second set of stimuli was created by pairing syllables that had prominence in their second syllable (i.e., vi-pa, su-tu, ke-ma). In each trial, a pair of syllables (e.g., pa-su or su-tu) was repeated for 14 seconds, with an interval of 200 ms between pairs of syllables. This resulted in a total of six different trials, each repeated twice, resulting in a total of 12 test trials.

In order to gain statistical power for the comparison of the two main conditions (iambs vs. trochees), pairs of non-adjacent syllables were not included in this second experiment. The test stimuli were synthesized with the acoustic parameters from Experiment 1, they were spondees with all the syllables having the same pitch (180 Hz) and phoneme duration (180 ms). The same pairs of syllables were used in the three experimental conditions, thus any difference between the experiments can only be due to the effect of the different streams infants were exposed to during the familiarization phase.

3.3 Procedure

Throughout the experiment, the infant was seated on the lap of the caregiver on a chair in the center of a booth. In front of the infant, a green lamp was fixed. On the left and right sides of the booth, yellow lamps were fixed at the same height as the green lamp. Loudspeakers were placed behind the yellow lamps.

The experimenter was located outside the testing booth. By looking at movements of the infant projected onto a video screen, the experimenter could control the presentation of the acoustic stimuli, and the flashing of the yellow and green lamps. Both the experimenter and the caregiver were "deaf" in regards to the stimuli being played in the testing booth.

The design of the experiment followed a modification of the head-turn preference paradigm (Saffran, Johnson, Aslin, and Newport, 1999). During the familiarization phase, the infant listened to one of the three familiarization streams from the experiment (i.e., prominence marked by pitch, prominence marked by duration, and a control stream). While the infant listened to the stream, the lights on the sides or on the center of the testing booth flashed, contingently upon the infant's looking behavior, and independently of the sound being played.

During the test phase, the infant was confronted with 12 trials (each trial with each of the 6 words repeated for 14 seconds). The trials started with the center light flashing, in order to attract the infant's attention. Once the infant looked at the central light, it was extinguished, and the light on either the left or right side of the booth started flashing. When infants fixated on the side light (defined as a 30° head turn towards the light), the sound started playing from the speaker behind the corresponding light. If the infant looked away for more than two seconds, the trial was stopped, and another trial started. In cases where the infant held his/her head towards the presentation side constantly, the entire trial was presented. The maximum looking time per trial could be the sum of several head turns, or a single head turn. The experiment lasted around 4 minutes.

The side and the order of the presentation of the stimuli were counterbalanced within each of the three experimental conditions. Infants were videotaped during the experiment, and their looking behavior was coded off-line by "blind" coders.

3.4 Results

The total looking times for the words with initial and final prominence were calculated for each infant. Infants familiarized with the stream varying in pitch showed during the test phase a preference to listen to pairs of syllables that during familiarization had high pitch in the first syllable, $t(15) = 2.584$, $p = 0.021$. When infants were familiarized to the stream varying in duration, no preference for sequences with longer syllables in initial or final position was found, $t(15) = 1.417$, $p = 0.177$. In the control condition, in which all syllables had the same pitch and duration, no preference for any type of syllable was found either, $t(15) = -1.064$, $p = 0.304$. A repeated measures general linear model was fitted to the data, with prominence during familiarization (syllable initial and syllable final) as a within subject factor and condition (pitch, duration, and flat) as a between subject factor. The model reveals no main effect of condition, $F(2, 45) = 2.092$, $p = 0.135$, no main effect of prominence, $F(1, 45) = 2.276$, $p = 0.138$, but a significant interaction between prominence and condition, $F(2, 45) = 3.811$, $p = 0.030$ (Figure 3). Therefore, infants' preference for different sequences of syllables during the test phase (prominence) was influenced by the stream they heard during familiarization (condition).

In short, the results showed that a bias to group linguistic stimuli trochaically when prominence is marked by pitch is already present in pre-verbal infants. When infants were familiarized with a sequence alternating in pitch, during the test phase they showed a preference to listen to linguistic units that were previously heard with initial prominence. In other words, they preferred sequences that during familiarization were coherent with the claims of the 'iambic-trochaic' law.

In contrast, infants who were familiarized with the sequence alternating in duration or with constant duration and pitch did not show any preference. Importantly, the stimuli during the test phase were presented with no prominence, thus any difference in infants' preferences can only be due to the different familiarization phases.

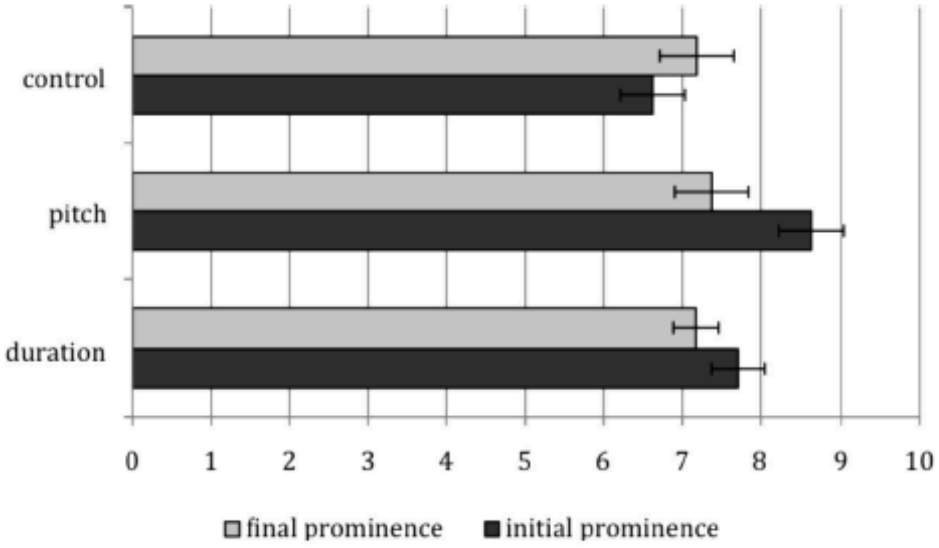

Figure 3. Mean looking times (seconds) and standard errors for the three experimental groups on the two sets of stimuli

4 Discussion and conclusions

In the two studies presented here, a similar set of stimuli was used to investigate adults' and infants' perceptual biases. Specifically, our studies focused on how prominence marked by either pitch or duration influenced infants' and adults' segmentation of speech sequences. The findings from the adult experiment were consistent with our predictions. Adults followed the principles of the iambic-trochaic law; they were better at remembering linguistic units that ended in a long element, or that started with a high-pitched element. These results add to the body of literature on the iambic-trochaic law, providing evidence that Italian speakers, like French and English speakers (Hay and Diehl, 2007), segment auditory streams following the iambic-trochaic law. This is also the first study to show that pitch, in addition to intensity, induces trochaic grouping in adults.

The findings from the infant study, in contrast, were not entirely consistent with the iambic-trochaic law. Infants showed trochaic biases for the stream varying in pitch, but showed no preferences for the stream varying in duration. It seems, therefore, that perceptual grouping based on pitch and perceptual grouping based on duration follow different developmental paths. Seven-month-old infants already show perceptual grouping similar to adults' for sequences alternating in pitch, but not for sequences alternating in duration.

Interestingly, these results support those of several other recent studies on the iambic-trochaic law. Studies with adults indicate that perceptual grouping by duration, instead of being guided uniquely by innate perceptual principles, might also be influenced by the properties of one's native language (Iversen, Patel, and Ohgushi, 2008). Iversen and colleagues (2008) compared the perception of a sequence of tones alternating in either intensity or duration by Japanese and English listeners. The two groups of listeners did not differ in their grouping of tones alternating in intensity, always placing the prominent element in initial position (trochee). However, the two groups differed in their grouping of stimuli alternating in duration. Specifically, English listeners tended to place the long stimuli in final position (grouping the sequences as iambs), while the Japanese on average did not show a preference for either iambs or trochees.

A similar pattern of results was found by Yoshida and colleagues (2010) with Japanese- and English-learning infants. In this study, infants' listening preferences for sequences of alternating long and short tones (trochees) or short and long tones (iambs) were investigated. No listening preferences were found for either the Japanese- or the English-learning participants at 5 months of age. However, English-learning 8-month-olds preferred to listen to sequences of iambs, but the Japanese infants did not show a preference either way – similar to the findings reported by Iversen et al. (2008) with adults. These results are in agreement with our findings that segmentation biases based on duration emerge from language experience.

In yet another study, similar conclusions were reached (Hay and Saffran, 2008). In this study, the iambic-trochaic bias was contrasted with distributional cues, and 9-month-old infants relied more strongly on prosody (trochees signaled by intensity and iambs signaled by duration) than on transitional probabilities when segmenting a continuous speech stream or a sequence of tones. However, at 6 months of age, trochaic biases using intensity are present, but no bias for segmentation based on duration is observed. This extensive set of studies supports our hypothesis that grouping biases based on duration emerge from language experience and are not yet present at 7 months of age, but emerge around 9 months of age.

However, we found trochaic biases for the pitch condition, and these results have implications for the different proposals put forward to explain early trochaic biases in infancy. The rhythmic-activation proposal claims that trochaic biases should be present in infants learning stress-timed languages such as English or Dutch, while no bias should be present in infants learning syllable-timed languages such as French or Italian (Nazzi et al., 2006). The existence of a trochaic bias in our pitch condition for Italian infants seems to be problematic for this proposal. It is possible therefore that the lack of a trochaic bias in French-learning infants might be due to the acoustic characteristics of the stimuli used in the different experiments (Höhle et al., 2009), in which both duration and pitch co-occur. Those cues often co-occur in natural languages, but owing to their different weights in different languages, descriptions of infants' perceptual biases should be more specific at the acoustic level.

Our results also do not support a proposal for innate trochaic bias without taking into account the acoustic features of the stimuli (Allen and Hawkins, 1978), or the existence of an innate iambic-trochaic bias. This is due to the fact that trochaic biases were observed in our experiment only for the stream alternating in pitch, while no iambic bias was present for the stream alternating in duration. It therefore appears that different developmental paths are present for grouping biases based on duration and pitch, with the former emerging later in development.

Developmental changes in the weighting of pitch and duration have been observed in other domains of linguistic processing. For instance, Seidl and colleagues (Seidl and Cristià, 2008) found that 4-month-old infants weight acoustic cues to signal phrasal boundaries differently from 6-month-olds. The authors found that at 4 months of age, pitch, duration, and intensity are used in conjunction for the segmentation of clauses, while at 6 months of age, pitch is used as the main cue for segmentation (Seidl, 2007). Importantly, these developmental changes were not captured by previous studies that failed to manipulate pitch and duration separately (Nazzi, Kemler Nelson, Jusczyk, and Jusczyk, 2000). Acoustic cues are also weighted differently in child-directed speech according to the age of the infant being talked to (Kitamura and Burnham, 2003), and the presence of different acoustic cues in infant-directed speech also influences infants' listening preferences differently throughout development (Panneton, Kitamura, Mattock, and Burnham, 2006). For instance, slower speech enhances 4-month-olds' attention to infant-directed speech, while it does not influence 8-month-olds' listening preferences (Panneton et al., 2006). Language experience also affects the weighting of cues such as duration at the word level, with sensitivity to vowel duration being highly dependent on the phonological weight of this cue in the infants' native language (Dietrich, Swingley, and Werker, 2007; Mugitani, Pons, Fais, Werker, and Amano, 2008). Further studies should investigate how attention to acoustic cues at different levels of linguistic processing changes throughout infancy, and how acoustic information such as pitch and duration at the segmental, word stress, clausal, and affective and pragmatic level are kept apart or interact to influence infants' early perceptual biases. Interestingly, the different weighting of different acoustic cues throughout infancy might in itself signal the linguistic units that warrant infants' attention at their current developmental stage.

In addition to different weighting of acoustic cues, it is also possible that perceptual grouping reflects infants' sensory abilities. Previous studies have shown that 7-month-old infants show pitch perception abilities similar to adults', being able to discriminate harmonic complexes from pitch categories differing by only 20% (e.g., 160 vs. 200 Hz, Clarkson and Clifton, 1985). Other studies seem to indicate that infants' abilities to discriminate variations in pitch are superior to adults' in some cases (Trainor and Trehub, 1992). It seems that the ability to detect variations of pitch is already present at birth (Nazzi, Floccia, and Bertoncini, 1998). The ability to discriminate changes in duration, however, seems to only be reached between the ages of 8 and 10 years (Elfenbein, Small, and Davis, 1993).

The question that remains is whether this early bias for grouping based on pitch observed in infants can still help them to learn about language-specific word order. We believe that it can: if the prominent element of phrases in the language is marked by pitch, infants can assume that the language is head-final; otherwise, that it is head-initial. This logic might be enough for infants to discover the ordering of words within phrases in their language, and adding another acoustic cue to signal word order might appear redundant. This proposal should be appealing to both generative and statistical approaches to language acquisition. Generative approaches might see this correlation either as a way to set universal switches during language acquisition or as a way to deviate from the default value just in case a signal indicates that it is necessary to do so. In a parameter-setting perspective, it is possible to argue that the parameter for head-directionality is initially set to head-initial, only changed to head-final in cases where prominence in the language is marked by duration. Statistical approaches, in contrast, might see this correlation as a way to organize the input into linguistically relevant units, or as a probabilistic cue aiding the acquisition of syntax. This bias might be directed towards universal acoustic cues, or it might instead be a general ability to develop sensitivity to the particular cues in the language being learned.

These perceptual biases can in turn be combined with other cues in the learning of language-specific word order. The unreliability of prosodic cues is not a problem if prosodic information is supplemented with other types of phonological cues and with distributional information (Christiansen, Allen, and Seidenberg, 1998; Hirsh-Pasek, Tucker, and Golinkoff, 1996). Recent studies have shown that infants' perceptual grouping is also influenced by the frequency of occurrence of the alternating elements in a stream, which also correlates with head-directionality within and across languages (Gervain, Nespor, Mazuka, Horie, and Mehler, 2008). All these cues, in combination with different perceptual biases, such as attention to adjacent repetitions, and biases for paying attention to the edges of linguistic constituents (Endress, Dehaene-Lambertz, and Mehler, 2007; Endress and Mehler, 2005; Gervain, Macagno, Cogoi, Peña, and Mehler, 2008) might be used in conjunction to discover language-specific word order regularities.

Independently of its use for syntactic bootstrapping, these biases for grouping of linguistic stimuli might be beneficial to both infants and adults. Humans have been shown to have a limited processing capacity (around seven chunks, Miller, 1956; or even fewer, Cowan, 2001), thus segmenting our sequences as three pairs of syllables instead of six individual syllables might decrease the processing load. Specifically, interpreting the sequence /pa/ /su/ /tu/ /ke/ /ma/ /vi/ as /pasu/ /tuke/ /mavi/ might improve memorization in a similar way as does segmenting the sequence BBCNBCCNN as three names of broadcasters (BBC, NBC, CNN). It has indeed been found that variations in pitch induce perceptual grouping of long speech sequences, such as phone numbers, and enhance memory performance (Frankish, 1995). Even pre-verbal untrained infants use chunking to increase their memory (Feigenson and Halberda, 2008). Importantly, these iambic-trochaic biases allow acoustic events to be segmented instantly without any previous experience with the stream, differently from classical statistical computations, which are highly experience dependent (Saffran, Aslin, and Newport, 1996). Once these events are found, further structural regularities can start to be extracted from the signal (Peña, Bonatti, Nespor, and Mehler, 2002).

In our pitch condition, infants are clearly segmenting the stream as three bisyllabic words, instead of into six individual syllables. In the duration condition, however, it is not clear what are the exact units infants are extracting from the stream. Infants might be extracting syllables, as a syllable-based segmentation strategy would propose, or bigger chunks of two, three, up to even six syllables. They are not segmenting the stream into three adjacent bisyllabic units, as this would probably result in a listening preference when those units are paired against units spanning a word boundary. A series of separate experiments would be necessary to distinguish the consistent units, if any, infants might be extracting from the stream. Our experiments show an asymmetry, at 7 months of age, in the segmentation of a speech stream when it alternates in duration or pitch, whereas adults show trochaic biases in a stream alternating in pitch, and iambic biases in a stream alternating in duration. Infants, in contrast, show perceptual biases only for bisyllabic segmentation in a stream alternating in pitch. The exact nature of the extraction of chunks from the stream alternating in duration cannot be identified at present, as it cannot be predicted from the iambic-trochaic law alone.

This study is the first step in investigating how different acoustic cues influence iambic and trochaic biases in very young infants. Further studies should use different syllable durations and pitch contours in the infant experiment. Specifically, while the increase in duration present in the adult experiment was enough to trigger iambic segmentation, longer durations than those we used might be necessary for the iambic bias to emerge in infants. In addition, it is not clear up to what extent variability at the segmental level influences the reliance on segmentation cues on the suprasegmental level. Most studies testing the iambic-trochaic law have focused on repetitions of a single isolated word, syllable, or tone. Hay and Diehl (2007), for instance, tested adults on several repetitions of the syllable /ga/ with variations in duration or intensity. Studies with infants used simplified streams as well. Trainor and Adams (2000) have used a single tone with variation in either intensity or duration in their stimuli, Yoshida et al. (2010) used a single tone with variation in duration, and Höhle and colleagues (2009) used several repetitions of the word /gaba/ with stress in the first or last syllable. In all these experiments, despite the simplicity of the stream, segmentation biases were observed. The hypothesis that less variation makes infants more sensitive to prosodic regularities is supported by studies showing that newborns are able to discriminate stress patterns when vowels do not change, but fail to do so when they vary (Sansavini, Bertoncini, and Giovanelli, 1997; Van Ooijen, Bertoncini, Sansavini, and Mehler, 1997).

Alternatively, it is possible that variability at the segmental level could help infants pay attention to prosodic cues. It has been shown that increased variability in a speech stream makes infants better at learning invariant structures, such as non-adjacent rules (Gómez, 2002). However, in our experiment, infants do not have to learn rules, but rather memorize the syllables in a speech stream. In other words, infants have to pay attention to both words and prosodic cues in order to extract smaller sized units. Most studies with infants at the same age as our group use no more than three different words in their familiarization stream (Saffran et al., 1996), exactly the number of words we used in our experiment. Therefore, considering infants' limited short-term memory (Feigenson and Halberda, 2008), increasing the number of words in our stream might result in failed segmentation, even if the invariant prosodic cues become more salient.

The present study adds to a current debate on the origins of trochaic biases observed in infants from different language backgrounds. It is pointed out that trochaic and iambic biases are highly dependent on the acoustic cues present in the speech stream, with iambic biases emerging from prominence marked by duration and trochaic biases emerging from prominence marked by pitch. The exact emergence of these two biases, however, was shown to follow different developmental pathways. Grouping biases based on pitch seem to emerge earlier in development than grouping biases based on duration. It is possible that perceptual grouping based on duration emerges with language experience, or instead depends on perceptual maturation, or attention to particular linguistic cues. This study sheds light on how acoustic markers of prominence might influence segmentation and memory for speech sequences by infants and adults. The exact way different perceptual mechanisms and biases interact and develop in early infancy, and help infants discover words, phrases, and the relation among these units in the apparently continuous speech stream is still an open question.

Acknowledgements

This work was supported by a grant to the second author from MICIT (Ministerio de Ciencia y Tecnología), CONICIT (Consejo Nacional para Investigaciones Científicas y Tecnológicas) of Costa Rica. We thank Marijana Sjekloća for help with recruiting and testing participants.

Note

1 A trial was rejected if looking time was less than 1 second. An infant was rejected if he had more than three rejected trials.

References

Allen, G., and Hawkins, S. (1978). The development of phonological rhythm. In A. Bell and J. Hooper (Eds.), *Syllables and segments* (pp. 173–185). Amsterdsam: North-Holland.

Bion, R. A. H., Höhle, B., and Schmitz, M. (2007). The role of prosody on the perception of word-order differences by 14-month-old German infants. *Proceedings of ICPhS 2007*, 1537–1540.

Bleakley, D. (1973). The effect of fundamental frequency variations on the perception of stress in German. *Phonetica*, 28, 42–59.

Bolton, T. (1894). Rhythm. *American Journal of Psychology*, 6, 145–238.

Christiansen, M., Allen, J., and Seidenberg, M. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13, 221–268.

Christophe, A., Gout, A., Peperkamp, S., and Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics*, 31, 585–598.

Christophe, A., Guasti, M. T., Nespor, M., and Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: The case of the head-complement parameter. *Developmental Science*, 6, 213–222.

Clarkson, M. G., and Clifton, R. K. (1985). Infant pitch perception: Evidence for responding to pitch categories and the missing fundamental. *Journal of the Acoustical Society of America*, 77, 1521–1528.

Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, 25(2), 257–271.

Cooper, G., and Meyer, L. (1960). *The rhythmic structure of music*. Chicago, IL: University of Chicago Press.

Cooper, W. E., and Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114.

Crystal, T., and House, A. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, 88, 101–112.

Cutler, A. (2005). Lexical stress. In D. B. Pisoni and R. E. Remez (Eds.), *The handbook of speech perception* (pp. 264–289). Oxford: Blackwell.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25(4), 385–400.

Cutler, A., and Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 113–121.

De Casper, A. J., and Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208, 1174–1176.

Dietrich, C., Swingley, D., and Werker, J. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences of the USA*, 104, 454–464.

Dogil, G. (1999). The phonetic manifestation of word stress in Lithuanian, Polish and German and Spanish. In H. van der Hulst (Ed.), *Word prosodic systems in the languages of Europe* (pp. 273–311). Berlin: Mouton de Gruyter.

Dupoux, E., Pallier, C., Sebastián-Gallés, N., and Mehler, J. (1997). A distressing "deafness" in French. *Journal of Memory and Language*, 36, 406–442.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., and Vreken, O. (1996). The Mbrola project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In H. T. Bunell and W. Isardi (Eds.), *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 1393–1396). Wilmington, DE: Alfred I. duPont Institute.

Elfenbein, J. L., Small, A. M., and Davis, J. M. (1993). Developmental patterns of duration discrimination. *Journal of Speech and Hearing Research*, 36, 842–849.

Endress, A., Dehaene-Lambertz, G., and Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition*, 105(3), 577–614.

Endress, A., and Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General*, 134(3), 406–419.

Feigenson, L., and Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 9926–9930.

Frankish, C. (1995). Intonation and auditory grouping in immediate serial recall. *Applied Cognitive Psychology*, 9(7), S5–S22.

Friederici, A., Friedrich, M., and Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology*, 17, 1208–1211.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., and Mehler, J. (2008). The neonate brain detects speech structure. *PNAS*, 105(37), 14222–14227.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., and Mehler, J. (2008). Bootstrapping word order in prelexical infants: A Japanese–Italian crosslinguistic study. *Cognitive Psychology*, 57, 56–74.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, 15, 103–113.

Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, 5, 431–436.

Gout, A., Christophe, A., and Morgan, J. (2004). Phonological phrase boundaries constrain lexical access: II. Infant data. *Journal of Memory and Language*, 51, 547–567.

Hay, J., and Diehl, R. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception and Psychophysics*, 69(1), 113–122.

Hay, J., and Saffran, J. (2008). Perceptual constraints on segmentation of speech and non-speech. Paper presented at the 33rd BUCLD, October 31–November 2, 2008, Boston, MA.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago, IL: University of Chicago Press.

Hirsh-Pasek, K., Tucker, M., and Golinkoff, R. (1996). Dynamic systems theory: Reinterpreting "Prosodic Bootstrapping" and its role in language acquisition. In J. Morgan and K. Demuth (Eds.), *Signal to syntax* (pp. 453–465). Mahwah, NJ: Erlbaum.

Höhle, B. (2002). *Der Einstieg in die Grammatik: Die Rolle der Phonologie/Syntax-Schnittstelle für Sprachverarbeitung und Spracherwerb*. Freie Universität Berlin.

Höhle, B., Bijeljac-Babicb, R., Weissenborn, J., and Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, 32(3), 262–274.

Houston, D., Jusczyk, P., Kuijpers, C., Coolen, R., and Cutler, A. (2000). Crosslanguage word segmentation by 9-month-olds. *Psychonomic Bulletin* and *Review*, 7, 504–509.

Iversen, J., Patel, A., and Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America*, 124(4), 2263–2271.

Johnson, E., and Seidl, A. (2008). A cross-linguistic perspective on the detection of juncture in speech. *Infancy*, 13(5), 440–455.

Jusczyk, P., Cutler, A., and Redanz, L. (1993). Preference for the predominant stress patterns of English words. *Child Development*, 64, 675–687.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., and Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24, 252–293.

Jusczyk, P., and Krumhansl, C. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance*, 19(3), 627–640.

Kitamura, C., and Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, 4(1), 85–110.

Kurby, C., and Zacks, J. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12, 72–79.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, 33, 451–454.

Llisterri, J., Machuca, M., de la Mota, C., Riera, M., and Rios, A. (2003). The perception of lexical stress in Spanish. In M.-J. Solé, D. Recasens, and J. Romero (Eds.), *Proceedings of the XVth International Conference of Phonetic Sciences* (Vol. 2, pp. 2023–2026). Barcelona: Causal Productions.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.

Morgan, J. L., and Demuth, K. D. (Eds.). (1996). *Signal to syntax: Bootstrapping from speech to grammar in early language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mugitani, R., Pons, F., Fais, L., Werker, J., and Amano, S. (2008). Perception of vowel length by Japanese- and English-learning infants. *Developmental Psychology*, 45(1), 236–247.

Nakatani, L. H., and Aston, C. H. (1978). Perceiving stress patterns of words in sentences. *Journal of the Acoustical Society of America*, 63(supplement 1), S55(A).

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1–11.

Nazzi, T., Floccia, C., and Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development*, 21(4), 779–784.

Nazzi, T., Iakimova, I., Bertoncini, J., Frédonie, S., and Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic difference. *Journal of Memory and Language*, 54, 283–299.

Nazzi, T., Kemler Nelson, D., Jusczyk, P., and Jusczyk, A. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy*, 1(1), 123–147.

Nespor, M., Guasti, M. T., and Christophe, A. (1996). Selecting word order: The Rhythmic Activation Principle. In U. Kleinhenz (Ed.), *Interfaces in phonology* (pp. 1–26). Berlin: Akademie Verlag.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., and Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue e Linguaggio*, No. 2, 139–168.

Nespor, M., and Vogel, I. (2008). *Prosodic phonology*. Berlin: Mouton de Gruyter. (1st edition 1986. Dordrecht: Foris.)

Panneton, R., Kitamura, C., Mattock, K., and Burnham, D. (2006). Slow speech enhances younger but not older infants' perception of vocal emotion. *Research in Human Development*, 3(1), 7–19.

Peña, M., Bonatti, L., Nespor, M., and Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298(5593), 604–607.

Potisuk, S., Gandour, J., and Harper, M. P. (1996). Acoustic correlates of stress in Thai. *Phonetica*, 53, 200–220.

Ramus, F., Hauser, M., Miller, C., Morris, D., and Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288, 340–351.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science*, 274, 1926–1928.

Saffran, J., Johnson, E., Aslin, R., and Newport, E. (1999). Statistical learning of tonal sequences by human infants and adults. *Cognition*, 70, 27–52.

Sansavini, A., Bertoncini, J., and Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, 33, 3–11.

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, 57, 24–48.

Seidl, A., and Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, 11(4), 596–606.

Singh, L., White, K. S., and Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development*, 4, 157–178.

Swallow, K., and Zacks, J. (2008). Sequences learned without awareness can orient attention during the perception of human activity. *Psychonomic Bulletin & Review*, 15(1), 116–122.

Swallow, K., Zacks, J., and Abrams, R. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, 138, 236–257.

Thiessen, E., and Saffran, J. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, 39, 706–716.

Trainor, L., and Adams, B. (2000). Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns. *Perception and Psychophysics*, 62(2), 333–340.

Trainor, L. J., and Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 394–402.

Van Ooijen, B., Bertoncini, J., Sansavini, A., and Mehler, J. (1997). Do weak syllables count for newborns? *Journal of the Acoustical Society of America*, 102, 3735–3741.

Weissenborn, J., and Höhle, B. (Eds.). (2001a). *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition* (Vol. 1). Amsterdam: John Benjamins.

Weissenborn, J., and Höhle, B. (Eds.). (2001b). *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition* (Vol. 2). Amsterdam: John Benjamins.

Williams, B. (1985). Pitch and duration in Welsh stress perception: The implications for intonation. *Journal of Phonetics*, 13, 381–406.

Woodrow, H. (1951). Time perception. In S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1224–1236). New York: Wiley.

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., and Werker, J. F. (2010). The development of perceptual grouping biases in infancy: A Japanese–English cross-linguistic study. *Cognition*, 115(2), 356–361.

Zacks, J., and Swallow, K. (2007). Event segmentation. *Current Directions in Psychological Science*, 16, 80–84.

Appendix B. The functional distinctions of consonants and vowels in 12 month-olds

Developmental Science

Developmental Science 14:6 (2011), pp 1445–1458 DOI: 10.1111/j.1467-7687.2011.01089.x

PAPER

Consonants and vowels: different roles in early language acquisition

Jean-Rémy Hochmann,[1] Silvia Benavides-Varela,[1] Marina Nespor[2] and Jacques Mehler[1]

1. *Cognitive Neuroscience Department, SISSA, International School for Advanced Studies, Trieste, Italy*
2. *Facoltà di Psicologia, University of Milano Bicocca, Italy*

Abstract

Language acquisition involves both acquiring a set of words (i.e. the lexicon) and learning the rules that combine them to form sentences (i.e. syntax). Here, we show that consonants are mainly involved in word processing, whereas vowels are favored for extracting and generalizing structural relations. We demonstrate that such a division of labor between consonants and vowels plays a role in language acquisition. In two very similar experimental paradigms, we show that 12-month-old infants rely more on the consonantal tier when identifying words (Experiment 1), but are better at extracting and generalizing repetition-based srtuctures over the vocalic tier (Experiment 2). These results indicate that infants are able to exploit the functional differences between consonants and vowels at an age when they start acquiring the lexicon, and suggest that basic speech categories are assigned to different learning mechanisms that sustain early language acquisition.

Introduction

To acquire language, infants need both to learn words and to extract and generalize structural regularities that play a role in learning grammar, e.g. syntax. For instance, when hearing the utterance 'the girl kicks the boy', we need more than just the meaning of each word to understand the whole sentence. The listener has to understand the relation between the verb and both the subject and the object. Learning words requires memorizing specific elements of the input (e.g. *girl*, *boy*) and representing them in a format that allows their recognition and distinction from other words (e.g. *boy* vs. *toy*), whereas learning regularities about syntactic structures implies the ability to extract relations between elements of the input (e.g. whether the verb precedes or follows the object) and generalize them to new sentences. Young language learners could thus profit from a (partial) 'division of labor', such that one speech category might preferentially support the acquisition of the lexicon, whereas another might be more dedicated to the identification of structural regularities, in particular those signaling relations between constituents. Here, we will present a series of observations that hint at different roles for consonants and vowels in language acquisition. Namely, we propose that consonants are more involved with word identification and encoding, because they are better suited than vowels for categorical perception. Vowels, in contrast, carry prosodic variations and provide cues to determine the boundaries and the organization of syntactic constituents (Nespor & Vogel, 1986; Selkirk, 1984). This functional difference between consonants and vowels constitutes the Consonant-Vowel hypothesis, hereafter referred to as the CV hypothesis (Nespor, Peña & Mehler, 2003).

The CV hypothesis

Part I: The consonantal bias in lexical processes

The role of consonants in the CV hypothesis originates from the observation that across languages, consonants allow more quality distinctions than vowels, where by quality we mean distinctions in terms of articulatory features.[1] With only a few exceptions, consonants are indeed cross-linguistically more numerous than vowels. For example, in Malay the proportion is 20C: 5V; in Italian 24C: 7V; in Hausa 32C: 5V; in Arabic 29C: 3V; in Igbo 27C: 5V; in Sindhi 46C: 10V (see Maddieson, 2008).

---

[1] There are systems where consonants may be distinguished for duration (e.g. Italian, *caro–carro*; [káro]-[kárːo]; 'dear' – 'cart') and systems where vowels may be distinguished because of suprasegmentals, such as nasality or tones (e.g. French, *beau–bon*; [bo]-[bõ]; 'beautiful' – 'good'). Also in these types of systems, by and large, consonant distinctions outnumber vowel distinctions.

Address for correspondence: Jean-Rémy Hochmann, SISSA, International School for Advanced Studies, Cognitive Neuroscience Department, Via Bonomea, 265, 34136, Trieste, Italy; e-mail: jr.hochmann@gmail.com

Most importantly, across languages, five vowel systems are the most common and most systems have over 20 consonants (Nespor *et al*., 2003). Because consonants are more numerous than vowels, they are relatively more informative for lexical distinctions.

A second observation motivating our belief that consonants have a predominant role to play in lexical learning is that, in Semitic languages such as Arabic and Hebrew, lexical roots are only represented on the consonantal tier (Berent, Vaknin & Marcus, 2007; McCarthy, 1985; Prunet, Beland & Adrissi, 2000). For example, in Arabic, the root *ktb* has the lexical meaning related to 'write'. The different vowels intervening between the consonants serve to form different words and word-forms that are related in meaning (e.g. *katib*: writer, *kataba*: he wrote, *kitab*: book, *maktaba*: library, etc.). In contrast, there is no documented language that has lexical roots based only on vowels. Furthermore, the case of consonantal lexical roots may be an extreme case of the situation observed in other languages, where consonants are in general more informative than vowels for lexical distinctions (see Keidel, Jenison, Kluender & Seidenberg, 2007, for the analysis of the French adult lexicon). In the Supplementary Material, we analyze the lexicon of infants for French and Italian, two languages that differ in the ratio of consonants and vowels (French has 19 consonants and 13 vowels; Italian has 24 consonants and seven vowels). For both languages, we verify that the sequence of consonants is more informative than the sequence of vowels for word identification. This suggests that the lexical role of consonants does not change as a function of the ratio of consonants and vowels.

Part II: The role of vowels in signaling syntactic organization

The CV hypothesis attributes a specific role to vowels for the acquisition of syntax. This hypothesis is based on three observations, which we develop below. First, vowels are the main carriers of prosodic information, including rhythm (Lehiste, 1970; Ramus, Nespor & Mehler, 1999). Second, pre-lexical infants and neonates are sensitive to rhythm and to prosodic phrase boundaries (Christophe, Mehler & Sebastián-Gallés, 2001; Christophe, Nespor, Dupoux, Guasti & van Ooyen, 2003; Nazzi, Bertoncini & Mehler, 1998; Ramus, Hauser, Miller, Morris & Mehler, 2000). Third, prosodic and rhythmic information provides cues that correlate with important morphosyntactic properties (Morgan & Demuth, 1996; Nespor, Shukla & Mehler, 2011; Nespor & Vogel, 1986; Selkirk, 1984). Syntax acquisition may thus start with inferences from prosodic cues carried by vowels.

In fact, vowels more than consonants can vary in terms of pitch, intensity and duration in relation to their sentential position. That is, vowels are more affected than consonants by prosody, which provides signals to syntactic constituency (Gleitman & Wanner, 1982; Morgan & Demuth, 1996; Nespor & Vogel, 2008; Selkirk, 1984). In particular, a phonological phrase boundary always coincides with a syntactic boundary. These boundaries are available to infants (Kemler Nelson, Hirsh-Pasek, Jusczyk & Wright-Cassidy, 1989; Gerken, Jusczyk & Mandel, 1994) as well as newborns (Christophe *et al*., 2001).

Moreover, certain syntactic properties correlate with prosodic cues across languages. For example, complements are marked mainly by pitch and intensity prominence in head final languages such as Turkish, and mainly by duration prominence in head initial languages such as French (Nespor, Shukla, van de Vijver, Avesani, Schraudolf & Donati, 2008). Those prosodic cues remain valid within languages that allow both types of constructions. For example, in German, the complement-head structures are mainly marked by an initial pitch and intensity prominence, while the head-complement structures are mainly marked by a final duration prominence (Nespor *et al*., 2008; Shukla & Nespor, 2010). Observing this and other correlations, prosodic bootstrapping theories (Morgan & Demuth, 1996) of language acquisition have proposed the existence of a bridge between prosodic and syntactic properties.

Finally, vowels carry a second type of rhythmic information. The proportion of time occupied by vowels in the speech input determines the rhythmic class of languages. Ramus *et al*. (1999) showed that vowels occupy about 45% of the speech stream in stress-timed languages (e.g. Dutch, English), about 50% in syllable-timed languages (e.g. French, Italian) and 55% in mora-timed languages (e.g. Japanese). Newborns can use this information to discriminate between two languages that belong to different rhythmic classes (Ramus *et al*., 2000). The rhythmic class to which a language belongs correlates with important morphosyntactic properties (Nazzi *et al*., 1998; Nespor *et al*., 2011; Ramus *et al*., 1999). In particular, the percentage of the speech stream that vowels occupy is indicative of the complexity of the syllabic repertoire of a given language. In turn, syllabic complexity correlates with the length of common words (Mehler & Nespor, 2004). In addition, typological studies have shown that languages with simple syllabic structures tend to be verb final, to use post-positions and have a rich case system (Donegan & Stampe, 1983; Gil, 1986; Fenk-Oczlon & Fenk, 2005; Nespor *et al*., 2011). Infants may therefore infer certain morphosyntactic properties of their language from the identification of its rhythmic class.

In sum, the prosodic – in particular rhythmic – properties of vowels provide infants with information about the syllabic repertoire, signal important syntactic boundaries, and provide cues to fundamental syntactic properties such as the relative order of heads and complements. These observations inspired the second part of the CV hypothesis: infants may focus on vowels when extracting structural information from their input, especially when learning about the relation between different sentence constituents.

Experimental evidence in favor of the CV hypothesis

A number of experimental results support the CV hypothesis. Infants are able to use statistical information, such as dips in transition probabilities (TPs) between syllables to identify word boundaries in a continuous speech stream (Saffran, Aslin & Newport, 1996). Newport and Aslin (2004) argued that adults could also compute TPs both between successive consonants and between successive vowels. However, assuming that participants use TPs to identify potential words, the CV hypothesis predicts that they should perform better at computing TPs over consonants than over vowels. In fact, in their experiments, Newport and Aslin (2004) allowed adjacent repetitions of the consonantal or vocalic frames to segment. Bonatti, Peña, Nespor and Mehler (2005) showed that, avoiding the adjacent repetition of both kinds of frames, adult participants are able to segment a continuous speech stream using the dependencies between consonants, but they fail to segment the stream using the dependencies between vowels. This suggests that participants can compute TPs over consonants but not over vowels. Moreover, Mehler, Peña, Nespor and Bonatti (2006) showed that when, within a single stream, TPs between consonants and TPs between vowels predict different segmentations, the statistics over consonants are favored. Thus, consonants appear to be a privileged category for discovering words in a continuous speech stream.

Word learning experiments in infants and toddlers further confirmed the advantage of consonants in encoding lexical items. Nazzi and colleagues (Nazzi, Floccia, Moquet & Butler, 2009) showed that in a word learning situation where 30-month-olds must ignore either a consonantal one-feature change or a vocalic one-feature change (e.g. match a /duk/ with either a /guk/ or a /dɔ̃k/), both French- and English-learning infants choose to neglect the vocalic change rather than the consonantal change. This preference was observed for word-initial (/guk/-/duk/-/dɔ̃k/), word-final (/pib/-/pid/-/pɛd/) and word-internal consonants (/gito/-/gipo/-/gupo/), and did not depend on an inability to process fine vocalic information. In agreement with these results, 16- to 20-month-old infants could acquire simultaneously two words differing only in one consonant, whereas they could not do so for minimal pairs differing in one vowel (Havy & Nazzi, 2009; Nazzi, 2005; Nazzi & Bertoncini, 2009). Furthermore, these findings are not restricted to one specific consonantal class (Nazzi *et al.*, 2009; Nazzi & New, 2007). Similar results were also obtained with 3-year-old French children (Havy, Bertoncini & Nazzi, 2011). Finally, even 4- and 5-year-old children, who are able to process fine phonetic information for both vowels and consonants, still show a consonantal bias for word learning (Havy, 2009).

The second prediction of the CV hypothesis is that infants should focus on vowels when extracting structural information. Acquiring the syntax of natural languages requires abstraction abilities. Chomsky (1957) showed that linguistic productions rely on structures that cannot be described by statistical dependencies between words (modeled as Markov chains or finite-state automatons). Rather, the description of syntax must contain hierarchical structures defined over abstract categories.

To show generalization abilities in infancy, researchers must provide evidence that infants transfer a response learned for a set of stimuli to a new set of stimuli. The representation that has been generalized can be inferred by looking at what is common and what differs between the two sets. From the perspective of the acquisition of structural syntactic information, we want to test infants' ability to generalize an abstract structure that cannot be reduced to statistical dependencies.

Recently, researchers have tackled this matter by using structures defined by the relative positions of identical elements (e.g. ABB, ABA). Indeed, since the seminal work of Marcus and colleagues (Marcus, Vijayan, Bandi Rao & Vishton, 1999) repetition-based structures have been extensively used to test abstract generalization abilities in infants and adults (see Endress, Dehaene-Lambertz & Mehler, 2007; Endress, Scholl & Mehler, 2005; Johnson, Fernandes, Frank, Kirkham, Marcus, Rabagliati & Slemmer, 2009; Kovács & Mehler, 2009b; Saffran, Pollak, Seibel & Shkolnik, 2007). Marcus and colleagues used the following structures: ABB (instantiated by words like *pukiki*, *mesasa*, etc.) and ABA (instantiated by words like *pukipu*, *mesame*, etc.). They showed that, after being habituated to exemplars of one of the structures, infants could discriminate between novel exemplars of both structures, suggesting that they had extracted and generalized the structures, rather than memorized a series of speech sequences.

There is a current debate about the precise mechanisms that underlie the extraction and generalization of repetition structures (Endress, Nespor & Mehler, 2009; Endress *et al.*, 2005; Marcus *et al.*, 1999). In the current paper, however, we will not discuss the mechanisms involved, but rather the constraints that may apply to these mechanisms. What matters to our present investigation of the respective roles of consonants and vowels in language acquisition is that the generalization of a repetition structure requires abstraction abilities that cannot be reduced to either memory or statistical dependencies between syllables or other constituents. Thus, in that respect, we take the generalization of a repetition structure as being comparable to the extraction of syntactic relations. Repetition structures can be viewed as the relation between two elements: infants display a generalization ability when they can recognize a repetition structure in novel syllable sequences that they have not encountered before (Kovács & Mehler, 2009b; Marcus *et al.*, 1999). The generalization of repetition structures can thus be used as a simplified model that might clarify how infants extract syntactic information.

The CV hypothesis – claiming that vowels are favored in the extraction of structural information – predicts that repetition-based structures should be easier to detect and generalize when they are implemented over vowels than over consonants. Experimental work corroborated this prediction for adults. Toro, Nespor, Mehler and Bonatti (2008a) showed that adult participants could easily learn the ABA regularity over vowels and generalize it to words using new vowels and consonants but respecting the same structure. In contrast, they were unable to learn the same regularity over consonants. Adults remain unable to generalize ABA over consonants even when vowel duration was reduced to one-third of the duration of consonants, while they could generalize ABA over barely audible vowels (Toro, Shukla, Nespor & Endress, 2008b). Thus, the reliance on vowels for extracting repetition-based regularities is not solely due to a major acoustic salience. Rather, vowels and consonants are involved in different types of processes, as suggested by the existence of a different neural substrate for each category (Caramazza, Chialant, Capasso & Miceli, 2000; Knobel & Caramazza, 2007).

Our study

In the experimental work presented below, we ask whether the documented functional difference between consonants and vowels can play a role in early steps of language acquisition. Evidence for the lexical role of consonants exists for participants older than 16 months (Havy & Nazzi, 2009), who already have a sizable receptive vocabulary (about one hundred and sixty words according to the French version of the MacArthur-Bates Communicative Development Inventory developed by S. Kern; personal communication). Moreover, evidence for the specialized role of vowels was reported for adults (Toro *et al.*, 2008a, 2008b). Recently, Pons and Toro (2010) suggested that 11-month-olds were already better at learning the AAB structure over vowels than over consonants. In a preferential looking paradigm, infants habituated to words respecting the AAB structure over vowels (e.g. *dabale*, *tolodi*, *tibilo*) could discriminate between novel words respecting the same structure (e.g. *nadato*, *lotoba*, *dilite*), and words that did not respect the AAB structure (e.g. *dutone*, *lanude*, *bitado*). In contrast, infants habituated to the AAB structure instantiated over consonants (e.g. *dadeno*, *lulabo*, *nunide*) did not discriminate novel words respecting the structure (e.g. *dedulo*, *lulina*, *nunobi*) from words that did not respect the structure (e.g. *dutani*, *litedo*, *bihune*). However, Pons and Toro (2010) used the same vowels and consonants in the test phase as in the familiarization phase. In particular, test and familiarization words shared the repeated vowels (e.g. *dabale* as a familiarization word, and *batalo* as a test word). Thus, an alternative explanation of Pons and Toro's results is that infants process mainly vowels and learned the repeated tokens (i.e. *aa*, *oo*, *uu*, *ii*, *ee*). They did not necessarily generalize the AAB structure, but may rather remember repeated vowel sequences. Showing generalization would require using novel vowels and consonants to form novel words in the test phase.

Taken together, the results reviewed above do not fully demonstrate the functional specialization of consonants and vowels in infancy. In fact, all the results could be explained by a switch in a general processing bias from vowels (at 11 months) to consonants (by 16 months) in infancy, while functional specialization would emerge only in adulthood. Thus, the aims of our study are, first, to assess whether the consonantal bias for lexical processes is already functional by 12 months, an age at which infants just start building the lexicon (Lock, 1980; McShane, 1979; Stager & Werker, 1997); second, to assess 12-month-olds' ability to generalize repetition structures over consonants and vowels. Moreover, by testing 12-month-olds in two very similar paradigms, which vary only in those details that differentiate a word learning experiment from a situation promoting the discovery of structural relations, we directly assess the CV hypothesis of a functional specialization of consonants and vowels.

Twelve-month-olds' word learning abilities are still immature (Lock, 1980; McShane, 1979; Stager & Werker, 1997). Infants begin to learn words in a very fast manner only towards the end of the second year of life (Carey & Bartlett, 1978; Golinkoff, Church Jacquet, Hirsh-Pasek & Nandakumar, 1996; Golinkoff, Hirsh-Pasek, Bailey & Wenger, 1992; Heibeck & Markman, 1987) after the emergence of a series of conceptual constraints that support word learning (Halberda, 2003; Hochmann, Endress & Mehler, 2010; Mervis, 1987; Markman, 1990; Markman & Hutchinson, 1984; Markman, Wasow & Hansen, 2003; Soja, Carey & Spelke, 1985). If the lexical role of consonants is already set before this stage, it can constrain and shape infants' vocabulary development. Similarly, even if numerous studies have shown that infants are sensitive to the boundaries of syntactic constituents (Gout, Christophe & Morgan, 2004; Soderstrom, Seidl, Kemler Nelson & Jusczyk, 2003), it remains unknown how much 12-month-olds know about the relation among these constituents. If we can establish that vowels are privileged for the extraction of such relational information, we may increase our understanding of the mechanisms present in syntax acquisition.

Experiment 1 tests the first part of the CV hypothesis (see Figure 1A). We ask whether infants rely more on consonants or on vowels when distinguishing among words. Adapting the paradigm developed by Kovács (2008) and Kovács & Mehler (2009b), we teach infants that one word predicts a toy's appearance on one side of the screen (e.g. *dudu*), while another word predicts a toy's appearance on the other side (e.g. *keke*). Infants are then presented with an ambiguous word, composed of the consonants of the former word and the vowels of the latter (e.g. *dede*) (or vice versa, e.g. *kuku*). On this occasion, no toy appears. Reliance on consonants would lead infants to search for the toy on the location pre-

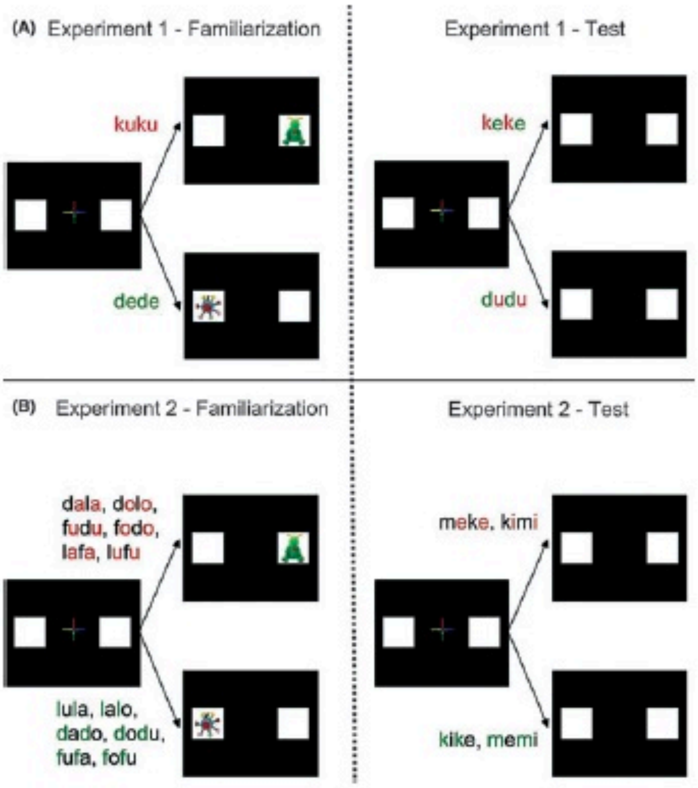

Figure 1 *Experimental paradigm for Experiments 1 (A) and 2 (B).*

dicted by the first word and reliance on vowels would lead infants to search for the toy on the location predicted by the second word. When forming memory representations of a novel word, the CV hypothesis predicts that infants should rely more on consonants than on vowels.

Experiment 2 tests the second part of the CV hypothesis (see Figure 1B), asking whether infants find it easier to learn and generalize regularities defined over vowels or over consonants. We again use the paradigm developed by Kovács (2008) and Kovács and Mehler (2009b). Monolingual 12-month-olds have trouble learning simultaneously two regularities in that paradigm, probably due to limitations in executive function (Kovács & Mehler, 2009b). Indeed, they just learn the simpler one (e.g. adjacent rather than non-adjacent syllable repetition, AAB vs. ABA; repetition rather than absence of repetition, AA vs. AB, ABA vs. ABC; Kovács, 2008). Thus, our paradigm allows us to test in a within-subject design which one of two structural relations is easier for infants to detect and generalize. In our experiment, we teach infants two regularities, consonant repetition and vowel repetition, in order to determine which is easier for 12-month-olds to learn. Each of six words containing a consonant repetition (i.e. *lula, lalo, dado, dodu, fufa* and *fofu*) was followed by a toy which appeared on one side of the screen, whereas each of six words containing a vowel repetition (i.e. *dala, dolo, fodo, fudu, lafa* and *lufu*) was followed by a toy which appeared on the other side of the screen. We then tested for generalization, asking where infants would search for the toy when hearing novel words respecting either the consonant repetition regularity (i.e. *kike* and *memi*) or the vowel repetition regularity (i.e. *meke* and *kimi*), using novel vowels and consonants that did not appear during the familiarization. These two regularities are strictly equivalent in terms of complexity (a simple repetition), varying only in the category that carries the repetition. The CV hypothesis predicts that the generalization of a repetition regularity should be easier if it is implemented over vowels than if it is implemented over consonants. Therefore, when learning only one regularity, infants should learn and generalize the vowel repetition rather than the consonant repetition.

Experiment 1

Participants

Twenty-six infants were included in the analysis; mean age: 12 months 6 days; range: 11 months 18 days–12 months 26 days. Five other infants participated in the study but were excluded due to fussiness (three) or experimental failure (two). Parents of the infants participating in the two experiments signed the informed consent explanation form before the experiments. The Ethics Committee of SISSA, where the experiments were conducted, approved the study design.

Stimuli

The words in Experiment 1 consisted of one syllable that duplicates. We used two consonants and two vowels to construct four nonsense words: *kuku*, *dede*, *keke* and *dudu*. The two consonants differ in three features, i.e. *k* is back, high and unvoiced, whereas *d* is anterior, coronal and voiced. The two vowels differ in three features, i.e. *e* is unrounded, mid and front, whereas *u* is rounded, high and back.

Two words sharing neither consonants nor vowels were used in the Familiarization (i.e. *kuku* and *dede*). The two remaining words were used in the Test. Words were synthesized with MBROLA (fr4) with a phoneme duration of 120 ms and a monotonous pitch of 200 Hz. There was no silent pause between two syllables within a word.

The visual stimuli were two pictures of colorful toys. Each appeared inside one of two white squares either on the left or on the right side of the screen. The toys loomed from 4 cm to 7 cm inside the squares for 2 s. The squares had a side-length of 8 cm, positioned at a distance of 13.5 cm. In the Familiarization, each toy was paired with one Familiarization word and one side.

Procedure

The procedure was adapted from Kovács and Mehler (2009b) and is presented in Figure 1. Stimuli were presented via an Apple Dual G5 computer running Psyscope X (http://psy.ck.sissa.it). Infants' gaze was recorded with a TOBII 1750 Eye-Tracker (Hofsten, Dahlstrom & Fredrikson, 2005).

The Familiarization phase consisted of 32 Familiarization trials. Familiarization trials started with a display of two white squares on the sides and a central attention-grabber. When the infant looked at the attention-grabber, either one of the two familiarization words was played in a pseudo-random order. We ensured that no word was repeated more than three times in a row. The animated attention-grabber was displayed until the offset of the word, in order to keep the infant's gaze in the middle of the screen. One second after the word offset, a toy appeared in one of the squares, contingent on the word: one word predicted the toy's appearance in one of the squares, while the other word predicted the toy's appearance in the other square. The pairing of the words with toy locations was counterbalanced across participants.

During test, infants were exposed to eight trials in a pseudo-random order. Test trials were similar to the Familiarization trials, except that infants heard words constituted by the consonants of one of the Familiarization words, and the vowels of the other. Thus if the Familiarization words were *dudu* and *keke*, the Test words were *dede* and *kuku*. If the Familiarization words were *dede* and *kuku*, the Test words were *dudu* and *keke*. No toy ever appeared in the test trials. Two seconds after the word onset, the next trial started. Test lists were pseudo-randomized so that the four first trials consisted in two trials for each test word.

Analysis

For the analysis, we divided the screen into three equal parts, left, middle and right. In each trial of the familiarization, we measured the proportion of infants anticipating the toy's appearance to the correct side. In the test, we measured each participant's first fixation, after hearing the new word and before the beginning of the next trial. Infants were coded as targeting either the consonant side or the vowel side. The vowel side was the one where the toy appeared after hearing the familiarization word that had the same vowels as the test word. For example, the vowel side for the test word *keke* was the side where during familiarization they learned to turn to after hearing the word *dede*, whereas the consonant side was where during familiarization they learned to turn to after hearing the word *kuku*. The consonant side for one of the two test words corresponded to the vowel side for the other test word. We also measured the infants' overall accuracy, based on the time spent fixating the consonant side and vowel side of the screen (Kovács & Mehler, 2009b; McMurray & Aslin, 2004). That is, for each trial, infants were scored as searching to the consonant side if the infant looked longer to the consonant side within the 2 s after hearing a new item and before the start of the next trial. Infants were scored as searching to the vowel side otherwise.

We computed difference scores: (#consonant looks – #vowel looks)/(#consonant looks + #vowel looks) for first and for overall accuracy, and computed a *t*-test to compare them to the chance level of 0. Positive differences in scores indicate that infants searched for the toys on the consonant side, while negative difference scores indicate infants searched for the toys on the vowel side.

The eight test trials might display extinction effects due to the absence of the puppets during this phase of the study. Thus, we also ran our analyses considering only the results of the first four test trials.

Results

The first fixation data for familiarization trials are presented in Figure 2. Infants anticipated one or the other side in 55% of the trials. We computed the proportion of infants showing a correct anticipatory look for each trial. A linear regression analysis showed no tendency for more correct anticipations in later familiarization trials, $\beta = .0011$, $R^2 = .0109$, $t(30) = .57$, $p > .57$.

The test results are presented in Figure 3. In the test phase, infants looked to the left or the right in 65% of the trials. Considering first fixations, infants' mean difference score was .23, which was significantly greater than 0, $t(25) = 2.1077$; $p = 045$; $d' = .41$. Seventeen infants displayed a positive difference score, six infants showed a negative difference score, and three infants a null difference score. A binomial test showed that significantly more infants displayed a positive difference score than a negative difference score, $p = .035$. Considering only the four first test trials, infants' mean difference score was .25, which was significantly greater than 0, $t(23) = 2.18$; $p = 039$; $d' = .45$. Fourteen infants displayed a positive difference score, six infants a negative difference score, and six infants a null difference score. The number of infants who displayed a positive difference score and the number of infants who displayed a negative difference score did not differ significantly, as evaluated by a binomial test, $p = 0.12$.

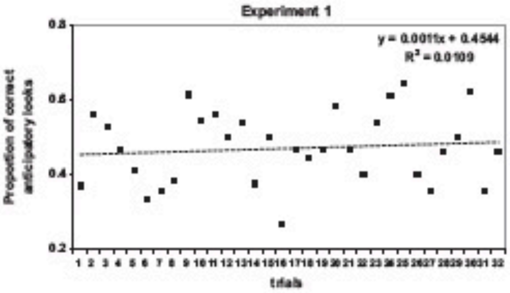

Figure 2 *Proportion of correct anticipatory looks for each familiarization trial in Experiment 1. The dotted line depicts the corresponding linear regression.*

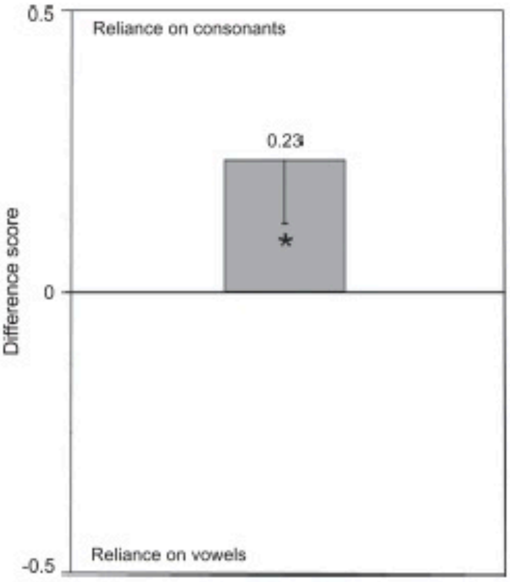

Figure 3 *Mean difference score for Experiment 1 considering the first fixations. Infants looked more at the side predicted by consonants. Error bars represent standard error.*

Considering the overall accuracy, infants' mean difference score was .13, which was not significantly different from 0; $t(25) = 1.33$; $p = .20$; $d' = .26$. Fifteen infants displayed a positive difference score, eight infants a negative difference score, and three infants a null difference score. A comparison of the number of infants who displayed a positive difference score and the number of infants who displayed a negative difference score was not significant, as evaluated by a binomial test, $p = 0.21$. Considering only the four first test trials, infants' mean difference score was .22, a non-significant trend in the predicted direction, $t(24) = 1.81$; $p = .072$; $d' = .36$. Fifteen infants showed a positive difference score, five infants a negative difference score, and five infants a null difference score. A binomial test showed that significantly more infants displayed a positive difference score than a negative difference score, $p = .041$.

Altogether, infants privileged the prediction made by consonants rather than that made by vowels.

Discussion

In this experiment, infants needed to learn that one word predicts a toy's appearance in one location, while another word predicts a different toy's appearance in another location. We further asked what prediction infants would make when presented with ambiguous words formed with the consonants of one of the previous words and the vowels of the other.

The observation of anticipatory looks in the familiarization did not show evidence of an increase in the number of correct anticipations. However, this absence of evidence should not be interpreted as infants' failure to learn the associations, as is shown by the test results. In fact, our paradigm is not designed to evaluate learning in the familiarization phase; rather, it is designed to assess the participants' performance in the test phase. In previous studies (Kovács, 2008; Kovács & Mehler, 2009b), the authors argued that due to limitations in executive function, most infants learned to predict the puppet's appearance for only one of the two screen locations. Similarly, in Experiment 1, most infants probably learned only to predict the puppet's appearance for only one of the two familiarization words. Thus, in all likelihood, infants' anticipations to the unlearned word, namely for 50% of the trials, should be assigned randomly to one or the other location. This predicts that data ought to be noisy during the familiarization phase.

Nevertheless, the observation of first fixations and the overall accuracy in the test phase suggests that infants consider two words sharing consonants as more similar than two words sharing vowels. Twelve-month-olds found *kuku* more similar to *keke* than to *dudu*.[2] A second interpretation is that infants relied solely on consonants when associating the familiarization words to either the side or the specific toy that appeared there. Our paradigm does not allow us to understand whether infants associated words to toys or to locations. In both cases, nevertheless, they needed to store at least one specific word in memory. When encoding specific words in memory, 12-month-old infants appear to give a higher weight to consonants than to vowels.

Experiment 2

In Experiment 2, we address the second part of the CV hypothesis, asking whether vowels are privileged for detecting and generalizing repetition structures. As in Experiment 1, infants heard speech sequences that predicted where toys would next appear. In Experiment 1, though, one word was associated with each location. In Experiment 2, six items were associated with each location. All six items associated with each location exemplified a common regularity, i.e. consonant repetition or vowel repetition.

Experiment 2, unlike Experiment 1, was not designed as a word learning experiment but as an experiment to elicit the generalization of a repetition structure. In that type of experiment, a variety of generalizations are possible from the limited set of examples that are provided to the participants. To avoid the position of specific phonemes providing a cue for generalizations, the same consonants and vowels were used to create the six consonant repetition and the six vowel repetition items used in the familiarization. Moreover, three puppets could appear in each location and were randomly paired with the items, so that the attention of infants could not be attracted by a particular puppet and an item associated with it. In the test phase, we asked whether infants could generalize these associations to novel items formed with novel consonants and vowels, instantiating the repetition structures. Predictions about the location of a toy's appearance in the test trials could be done only if one focused on the identity relation between consonants or vowels. Due to limitations in executive function, monolingual 12-month-olds usually learn to generalize only the easier structure when tested with this paradigm. We thus ask whether vowel repetition or consonant repetition is easier for 12-month-olds to generalize.

Participants

Twenty-four infants were included in the analysis; mean age: 12 months 4 days; range 11 months 26 days–12 months 28 days. Six other infants participated in the study but were excluded due to fussiness (three) or experimental failure (three).

Stimuli

The stimuli in Experiment 2 were bisyllabic items. These items could have either repeated consonants or repeated vowels. For the familiarization, six items containing a consonant repetition (*lula*, *lalo*, *dado*, *dodu*, *fufa* and *fofu*) and six items containing a vowel repetition were created (*dala*, *dolo*, *fodo*, *fudu*, *lafa* and *lufu*). The same three consonants and three vowels were used to generate both sets of items. For the test, four novel items were generated with novel consonants and novel vowels. Two test items had a consonant repetition (*kike* and *memi*) and two test items had a vowel repetition (*meke* and *kimi*). Items were synthesized with MBROLA (fr4) with phoneme durations of 120 ms and a monotonous pitch of 200 Hz. There was no silent pause between two syllables within an item.

Visual stimuli were three pictures of colorful toys. These appeared inside one of two white squares on the left or right side of the screen, as for Experiment 1. All three toys could appear after any familiarization item in the window predicted by that item.

Procedure

The procedure was identical to that of Experiment 1, except for the speech items infants heard (see Figure 1) and the toys they saw. In Experiment 2, a consonant repetition predicted the toys' appearance in one of the squares, while a vowel repetition predicted the toys' appearance in the other square. The pairing of structures with toy locations was counterbalanced across participants. Test lists were pseudo-randomized so that the four first and four last trials consisted of two trials for each structure. The structure of the first test item was vowel repetition for half the participants and consonant repetition for the other half.

---

[2] The Nazzi group has reported that, in older infants and children, the C-bias may vary depending on the actual contrasts used, even though it applies to most consonantal contrasts (Havy & Nazzi, 2009; Nazzi, 2005; Nazzi & Bertoncini, 2009; Nazzi *et al.*, 2009). In our experimental paradigm, infants were more likely to learn to predict the locations of the toys' appearances if the two novel words were easy to discriminate. For this reason, we chose words that differ by several rather than only one feature. Thus, we chose to use three-feature contrasts, both for consonants and for vowels. Further studies may ask whether one-feature contrasts yield similar results, particularly contrasts involving manner of articulation which did not vary in our experiment.

Analysis

The analysis was similar to that of Experiment 1, except that each fixation was coded as correct or incorrect, according to the structure that preceded infants' response, i.e. consonant repetition or vowel repetition. Independently for each structure, we computed difference scores: (#correct looks – #incorrect looks)/(#correct looks + #incorrect looks) for first fixations and for overall accuracy, and computed a *t*-test to compare them to the chance level of 0. Significantly positive difference scores would indicate that infants learned and generalized the structure.

Because the first four test trials contained only two test trials for each structure, unlike Experiment 1, we did not analyze separately these trials in Experiment 2.

Results

Familiarization results are presented in Figure 4. Infants anticipated to one or the other side in 56% of the trials. We computed the proportion of infants showing a correct anticipatory look for each trial. Qualitatively, the proportion of correct anticipatory looks increased during familiarization for the vowel repetition and decreased for the consonant repetition. A linear regression analysis yielded marginally significant results for the consonant repetition, $\beta = -.0159$, $R^2 = .21$, $t(14) = -1.92$, $p = .075$, and a non-significant trend for the vowel repetition, $\beta = .0046$, $R^2 = .15$, $t(14) = 1.60$, $p = .133$.

The test results are presented in Figure 5. Infants looked to the left or the right in 62% of vowel repetition test trials, and in 61% of the consonant repetition test trials. Two infants did not provide data in the vowel repetition tests, so that 24 infants were included in the analysis of the consonant repetition tests and only 22 in the analysis of the vowel repetition tests.

Paired *t*-test showed that infants obtained significantly higher difference scores for the vowel repetition than for the consonant repetition, considering the first fixations, $t(21) = 4.56$; $p < .0002$; d′ = .97; and the overall accuracy, $t(21) = 4.29$; $p < .0004$; d′ = .91.

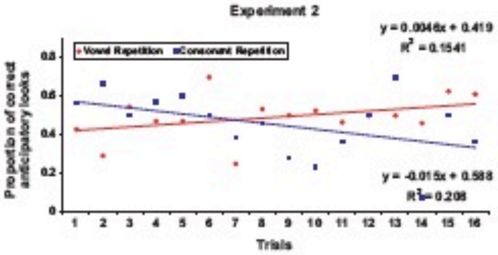

Figure 4 *Proportion of correct anticipatory looks for each familiarization trial in Experiment 2. The dotted lines depict the linear regression for vowel repetition (red) and for consonant repetitions (blue), respectively.*

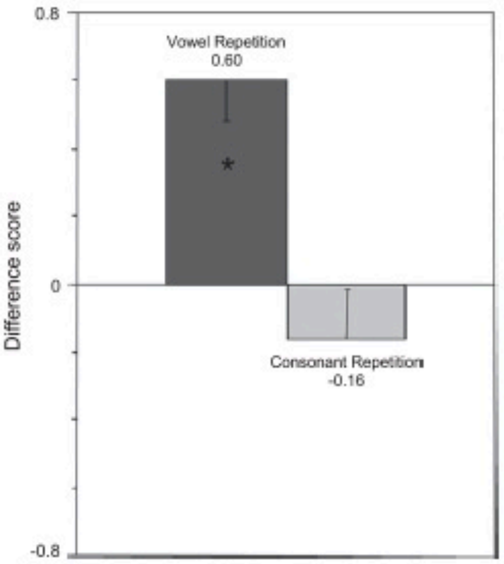

Figure 5 *Mean difference scores for Experiment 2 considering the first fixations. Infants looked more at the correct side predicted by the regularity for the vowel repetition, but not for the consonant repetition. Error bars represent standard errors.*

Considering first fixations in vowel repetition test trials, infants' mean difference score was .60 for the vowel repetition, which was significantly greater than 0, $t(21) = 4.92$; $p < .0001$; d′ = 1.05. Seventeen infants displayed a positive difference score, two infants a negative difference score, and three a null difference score. A binomial test showed that significantly more infants displayed a positive difference score than a negative difference score, $p < .001$. For the consonant repetition, infants' mean difference score was –0.16, which did not differ from chance, $t(23) = -1.06$; $p = .30$; d′ = .22. Eight infants displayed a positive difference score, 11 infants a negative difference score, and five a null difference score. This distribution did not differ from chance, $p = .144$.

Congruently, considering overall accuracy, infants' mean difference score was .53 for the vowel repetition, which was significantly greater than 0; $t(21) = 5.23$; $p < .0001$; d′ = 1.11. Fifteen infants displayed a positive difference score, one infant a negative difference score, and six a null difference score. A binomial test showed that significantly more infants displayed a positive difference score than a negative difference score, $p < .001$. Infants' mean difference score was –.07 for the consonant repetition, which did not significantly differ from chance, $t(23) = -.52$; $p = .61$; d′ = .11. Six infants displayed a positive difference score, nine infants a negative difference score, and nine a null difference score. This distribution did not differ from chance, $p = .61$.

Discussion

The CV hypothesis predicts that generalizing repetition structures should be easier over vowels than over consonants. Experiment 2 directly tested this claim by confronting infants with two competing regularities. Given the limited cognitive capacities of 12-month-olds (Kovács & Mehler, 2009a, 2009b), participants were expected to learn and generalize only one of the patterns. The results suggest that infants learned the association between the vowel repetition regularity and the predicted location of a toy's appearance. Moreover, they could extend this association to new, never heard, items with vowel repetition formed with novel consonants and vowels, thus showing that infants extracted an abstract property from the familiarization and represent an abstract vowel repetition structure. In contrast, they showed no evidence of learning and generalizing the consonant repetition structure. Thus, the same structure (i.e. a repetition[3]) is easier for 12-month-old infants to learn over vowels than over consonants. Vowels rather than consonants appear to be a privileged category for extracting and generalizing abstract structures.

General discussion

The CV hypothesis claims that infants rely more on consonants for word learning and recognition, but privilege vowels for extracting structural syntactic information. In two experiments, infants saw toys appearing in two distinct locations after hearing specific speech items. In Experiment 1, infants had to learn two different words, each associated with a specific location. Thereafter, in the test, infants relied more on consonants than on vowels in their search for the location of the toy. In Experiment 2, infants could generalize a repetition structure implemented over vowels, but not that implemented over consonants. Notably, Experiments 1 and 2 were designed in a similar way, and varied only in the details that should differentiate a word learning experiment from an experiment that evaluates the generalization of a structure. In Experiment 1, in addition to its location, each of the two toys was associated with a specific word, so that these could be interpreted as the names of the respective toys. In Experiment 2, the same consonants and vowels were used to implement both structures, the vowel repetitions and the consonant repetitions, in order to get infants to focus on the relation between segments rather than on the specific phonemes to predict the toy's location. The results allow us to conclude that consonants are privileged for lexical processes, whereas vowels better support the generalization of repetition structures. Thus, we show that functional differences between consonants and vowels are already playing a significant role by the end of the first year of life.

CV hypothesis, Part I: Consonants to build the lexicon

The end of the first year of life coincides with the time when infants begin to develop their vocabulary. According to the MacArthur-Bates Communicative Development Inventory (CDI; Dale & Fenson, 1996) questionnaire studies, infants of that age can understand about 80 words. Fenson, Dale, Reznick, Bates, Thal and Pethick (1994) evaluated that 12-month-olds learn about two words a week, a rate that improves in the following months yielding a 6-year-old vocabulary of about 10,000 words (Bloom & Markson, 1998; Miller, 1996). Thus, given that the consonantal bias for lexical acquisition appears to be in place at 12 months of age, it is likely to play a role in the acquisition of more than 99% of a child's vocabulary.

In the Supplementary Material, we show that the words in French and Italian infants' vocabulary can be better discriminated on the basis of the information carried by consonants than of that carried by vowels. In fact, Keidel *et al.* (2007) attributed the origin of the lexical role of consonants to the comparative distribution of information carried by vowels and consonants in the lexicon, which, in their view, would result from the larger number of consonants attested in the great majority of languages. However, the consonantal superiority for lexical processes is found in languages in which consonants largely outnumber vowels, as well as in languages in which the numbers of consonants and vowels are balanced (see Mehler *et al.*, 2006; Toro *et al.*, 2008a; and Bonatti, Peña, Nespor & Mehler, 2007, for Italian and French). For instance, Cutler, Sebastián-Gallés, Soler-Vilageliu and van Ooijen (2000) showed that, when asked to change one phoneme to turn a non-word (e.g. *kebra*) into a known word, participants more often altered the vowel (thus generating *cobra*) than the consonant (generating *zebra*). Their results were as significant for speakers of Spanish as for speakers of Dutch. Spanish has many more consonants than vowels, whereas Dutch has a similar number of consonants and vowels. Furthermore, in reading tasks, consonants appear to be privileged for lexical access in various languages such as French (New, Araujo & Nazzi, 2008), English (Berent & Perfetti, 1995; Lee, Rayner & Pollastek, 2001) and Spanish (Carreiras, Gillon-Dowens, Vergara & Perea, 2009). If the lexical statistics hypothesis fully explained the specialization of consonants for lexical access, a larger effect would be expected in languages that have many

---

[3] Note that, in our stimuli, an alternative description of the vowel repetition structure would be a *non*-repetition of consonants. Indeed, rather than generalizing the repetition of the vowel, generalizing the pattern in which the first and second consonants should differ would yield similar results. However, we consider this possibility unlikely, as Kovács (2008) showed how both 7- and 12-month-olds found it easier to learn and generalize a syllable repetition regularity rather than a non-repetition regularity.

more consonants than vowels. In contrast, we interpret the distribution of information in consonants and vowels as a consequence of the consonantal bias for lexical acquisition (see Bonatti *et al.*, 2007).

The functional differences between consonants and vowels may be innate aspects of humans' ability to acquire language, or originate from the fact that different processes have different requirements, and may consequently rely on different categories. Specifically, lexical memory may require more stable, thus reliable, categories that allow the learner to identify words and distinguish each word from other lexical entries. The lexical role of consonants may therefore be due to the categorical mode in which consonants are perceived. In fact, speech perception abilities change in the course of the first year of life. Infants are initially sensitive to all phonemic contrasts that can be found in the languages of the world (Jusczyk, 1997; Mehler & Dupoux, 1994). By 6 months of age, however, infants display a perceptual magnet effect for vowel perception, which is due to the formation of a prototype for each of the vowels of the language of exposure (Kuhl, 1991; Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992). By the end of the first year of life, infants converge to the consonantal categories of the language of exposure, and have lost some sensitivity to non-native consonantal contrasts (Werker & Tees, 1984). Even in adulthood, consonants are perceived more categorically than vowels. Indeed, in specific experimental conditions, when adult participants would fail to detect a within-category consonantal variation, they still detect a within-category vocalic variation (Fry, Abramson, Eimas & Liberman, 1962; Pisoni, 1973). As a consequence, vowel variations due to sentence prosody as well as speaker and dialectal accents may hinder word recognition to a greater extent than consonant variations. Therefore, lexical distinctions should be better instantiated by consonantal contrasts than by vocalic contrasts. This view predicts that the formation of consonantal phonological categories may be necessary for the lexical role of consonants to emerge.

CV hypothesis, Part II: Generalization of structural regularities over vowels

Inspired by the prosodic bootstrapping accounts of syntax acquisition (Morgan & Demuth, 1996), the CV hypothesis predicted that abstract structures should be easier to generalize over vowels than over consonants. In Experiment 2, we show that – in a within-subject design – 12-month-old infants are better at extracting a repetition-based structure over vowels than over consonants. Moreover, ours is the first experiment to show that infants can generalize the vowel repetition structure to completely novel words, formed with vowels and consonants that did not appear in the familiarization. This result indicates a special role of vowels in supporting the extraction and generalization of structures that cannot be reduced to statistical dependencies between constituents (e.g. syllables).

The origin of the specialization of vowels remains unknown, as does the origin of the consonantal bias in word learning. However, in the case of vowels, it is even harder to come up with a scenario where infants would have learned that vowels are more informative for specifying identity relations. Alternatively, the notion that vowels carry structural information may constitute an innate aspect of humans' ability to acquire language.

Consequently, 12-month-old and possibly younger infants ought to be capable of extracting the structural information carried by vowels, including prosodic information that informs syntax (Nespor & Vogel, 1986; Selkirk, 1984). For example, they may quickly learn and generalize prominence alternations signaled either by pitch (Bion, Benavides Varela & Nespor, in press) or by duration, which may allow them to learn the order of heads and complements in their language (Nespor *et al.*, 2008).

Obviously, we are not claiming that all linguistic regularities are acquired through the detection and generalization of repetition patterns. However, the acquisition of natural syntactic constituent structures, similarly to the identification of repetition structures, requires the ability to generalize structural relations that cannot be reduced to a statistical regularity and thus require computations that go beyond the reach of memory and statistical computations (Chomsky, 1957; Marcus, 1998). Such computations may involve the formation of algebraic rules (Marcus *et al.*, 1999), or the trigger of language-specific learning mechanisms. The precise characterization of the mechanisms involved will require further studies. However, here we have shown that these mechanisms are constrained in the speech domain, and particularly rely more on the vocalic tier than on the consonantal tier, suggesting that these mechanisms have evolved to suit the needs of human speech processing and language acquisition.

Conclusion

Here we document that the functional differences between consonants and vowels are already available to 12-month-old infants. The functional specialization of consonants is present early enough to shape the acquisition of most of the words a child will learn (possibly with the exception of a few very salient words learned very early on[4]). Furthermore, even though infants as young as 6 months are able to detect prosodic cues

---

[4] Note, however, that cross-linguistically, the words that indicate the mother and the father, two of the first words infants acquire, tend to contrast in their consonants but not in their vowels (e.g. *mamma/papa* in Italian; *anya/apa* in Hungarian; *aai/baba* in Marathi; *māma/bàba* in Mandarin Chinese; Jakobson, 1960; Murdoch, 1959). Thus the lexical role of consonants may already be set when these words are learned.

marking syntactic phrase boundaries (Gout *et al.*, 2004; Soderstrom *et al.*, 2003), it is less clear how they learn about the relation among them. Our results suggest that the structural information carried by vowels is available at least to 12-month-olds and may guide them in this task.

Infants may know or have a mechanism to identify which category better serves the lexicon and which the extraction of structural regularities. We proposed that the distributional and physical properties of consonants make them more reliable for word recognition and lexical distinctions, thus favoring them in word learning processes. Vowels, instead, carry prosody that marks more abstract constituents and provides the learner with information about structural relations. Future studies will explore the link between these specific properties of consonants and vowels and their specializations for language acquisition. These complementary constraints on the processing of consonants and vowels may allow the language acquisition device to work simultaneously on the extraction of regularities and on the acquisition of lexical items.

Acknowledgements

This research has been supported by McDonnell Foundation Grant No. 21002089 and by a grant to the second author from MICIT (Ministerio de Ciencia y Tecnología), CONICIT (Consejo Nacional para Investigaciones Científicas y Tecnológicas) of Costa Rica. We thank Francesca Gandolfo and Marijana Sjekloča for help recruiting participants, and Liuba Papeo and Alan Langus for valuable comments on a previous version of the manuscript.

References

Berent, I., & Perfetti, C.A. (1995). A rose is a reez: the two cycles model of phonology assembly in reading English. *Psychological Review*, **102**, 146–184.

Berent, I., Vaknin, V., & Marcus, G.F. (2007). Roots, stems, and the universality of lexical representations: evidence from Hebrew. *Cognition*, **104**, 254–286.

Bion, R.H, Benavides Varela, S., & Nespor, M. (in press). Acoustic markers of prominence influence infants' and adults' memory for speech sequences. *Language and Speech.*

Bloom, P., & Markson, L. (1998). Capacities underlying word learning. *Trends in Cognitive Sciences*, **2**, 67–73.

Bonatti, L.L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychological Science*, **16**, 451–459.

Bonatti, L.L., Peña, M., Nespor, M., & Mehler, J. (2007). On consonants, vowels, chickens, and eggs. *Psychological Science*, **18**, 924–925.

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature*, **403**, 428–430.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, **15**, 17–29.

Carreiras, M., Gillon-Dowens, M., Vergara, M., & Perea, M. (2009). Are vowels and consonants processed differently? Event-related potential evidence with a delayed letter paradigm. *Journal of Cognitive Neuroscience*, **21**, 275–288.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Infants' sensitivity for prosodic boundary cues. *Infancy*, **2**, 385–394.

Christophe, A., Nespor, M., Dupoux, E., Guasti, M.-T., & van Ooyen, B. (2003). Reflections on prosodic bootstrapping: its role for lexical and syntactic acquisition. *Developmental Science*, **6**, 213–222.

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: cross-linguistic comparisons. *Memory and Cognition*, **28**, 746–755.

Dale, P.S., & Fenson, L. (1996). Lexical development norms for young children. *Behavioral Research Methods, Instruments, & Computers*, **28**, 125–127.

Donegan, P., & Stampe, D. (1983). Rhythm and the holistic organization of language structure. *Papers from the Annual Regional Meeting, Chicago Linguistic Society: Parasession on the Interplay of Phonology, Morphology and Syntax* (pp. 337–353).

Endress, A.D., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition*, **105**, 577–614.

Endress, A.D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, **13**, 348–353.

Endress, A.D., Scholl, B.J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology, General*, **134**, 406–419.

Fenk-Oczlon, G., & Fenk, A. (2005). Crosslinguistic correlations between size of syllables, number of cases, and adposition order. In Gertraud Fenk-Oczlon & Christian Winkler (Eds.), *Sprache und Natürlichkeit: Gedenkband für Willi Mayerthaler* (pp. 75–86). Tübingen: Gunther Narr.

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., & Pethick, J. (1994). Variability in early communication development. *Monographs of the Society for Research in Child Development*, **59** (5, Serial No. 242).

Fry, D.B., Abramson, A.S., Eimas, P.D., & Liberman, A.M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, **5**, 171–189.

Gerken, L.A., Jusczyk, P.W., & Mandel, D.R. (1994). When prosody fails to cue syntactic structure: nine-month-olds' sensitivity to phonological vs. syntactic phrases. *Cognition*, **51**, 237–265.

Gil, D. (1986). A prosodic typology of language. *Folia Linguistica*, **20**, 165–231.

Gleitman, L.R., & Wanner, E. (1982). The state of the state of the art. In E. Wanner & L.R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3–48). Cambridge: Cambridge University Press.

Golinkoff, R.M., Church Jacquet, R., Hirsh-Pasek, K., & Nandakumar, R. (1996). Lexical principles may underlie the learning of verbs. *Child Development*, **67**, 3101–3119.

Golinkoff, R.M., Hirsh-Pasek, K., Bailey, L.M., & Wenger, N.R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, **28**, 99–108.

Gout, A., Christophe, A., & Morgan, J. (2004). Phonological phrase boundaries constrain lexical access: II. Infant data. *Journal of Memory and Language*, **51**, 547–567.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, **87**, B23–B34.

Havy, M. (2009). Spécificité phonétique et acquisition lexicale en situation de développement normal et atypique. Thèse de Doctorat en Psychologie Cognitive, Université Paris Descartes, soutenance publique: 21 décembre.

Havy, M., Bertoncini, J., & Nazzi, T. (2011). Word learning and phonetic processing in preschool age children. *Journal of Experimental Child Psychology*, **108**, 25–43.

Havy, M., & Nazzi, T. (2009). Better processing of consonantal over vocalic information in word learning at 16 months of age. *Infancy*, **14**, 439–456.

Heibeck, T.H., & Markman, E.M. (1987). Word learning in children: an examination of fast mapping. *Child Development*, **58**, 1021–1034.

Hochmann, J.-R., Endress, A.D., & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition*, **115**, 444–457.

Hofsten, C., Dahlstrom, E., & Fredrikson, Y. (2005). 12-month-old infants' perception of attention direction in static video images. *Infancy*, **8**, 217–231.

Jakobson, R. (1960). Why mama and papa? In B. Kaplan & S. Wapner (Eds.), *Perspectives in psychological theory: Essays in honor of Heinz Werner* (pp. 124–134). New York: International Universities Press.

Johnson, S.P., Fernandes, K.J., Frank, M.C., Kirkham, N.Z., Marcus, G.F., Rabagliati, H., & Slemmer, J.A. (2009). Abstract rule learning for visual sequences in 8- and 11-month-olds. *Infancy*, **14**, 2–18.

Jusczyk, P.W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Keidel, J.S., Jenison, R.L., Kluender, K.R., & Seidenberg, M.S. (2007). Does grammar constrain statistical learning? Commentary on Bonatti et al. (2005). *Psychological Science*, **18**, 922–923.

Kemler Nelson, D.G., Hirsh-Pasek, K., Jusczyk, P.W., & Wright-Cassidy, K. (1989). How prosodic cues in motherese might assist language learning. *Journal of Child Language*, **16**, 55–68.

Knobel, M., & Caramazza, A. (2007). Evaluating computational models in cognitive neuropsychology: the case from the consonant/vowel distinction. *Brain and Language*, **100**, 95–100.

Kovács, Á.M. (2008). Learning two languages simultaneously: mechanisms recruited for dealing with a mixed linguistic input. SISSA, Trieste. PhD dissertation.

Kovács, Á.M., & Mehler, J. (2009a). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences, USA*, **106** (2), 6556–6560.

Kovács, Á.M., & Mehler, J. (2009b). Flexible learning of multiple speech structures in bilingual infants. *Science*, **325**, 611–612.

Kuhl, P.K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, **50**, 93–107.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, **255**, 606–608.

Lee, H.W., Rayner, K., & Pollastek, A. (2001). The relative contribution of consonants and vowels to word identification during reading. *Journal of Memory and Language*, **44**, 189–205.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lock, A. (1980). *The guided reinvention of language*. London: Academic Press.

McCarthy, J.J. (1985). *Formal problems in semitic phonology and morphology*. New York and Oxford: Garland.

McMurray, B., & Aslin, R.N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy*, **6**, 203–229.

McShane, J. (1979). The development of naming. *Linguistics*, **17**, 879–905.

Maddieson, I. (2008). Consonant–vowel ratio. In M. Haspelmath, M.S. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library, ch. 3. Available online at: http://wals.info/feature/3

Marcus, G.F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, **37**, 243–282.

Marcus, G.F., Vijayan, S., Bandi Rao, S., & Vishton, P.M. (1999). Rule-learning in seven-month-old infants. *Science*, **283**, 77–80.

Markman, E.M. (1990). Constraints children place on word meanings. *Cognitive Science*, **14**, 57–77.

Markman, E.M., & Hutchinson, J.E. (1984). Children's sensitivity to constraints on word meaning: taxonomic vs. thematic relations. *Cognitive Psychology*, **16**, 1–27.

Markman, E.M., Wasow, J.L., & Hansen, M.B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, **47**, 241–275.

Mehler, J., & Dupoux, E. (1994). *What infants know*. Cambridge, MA: Blackwell.

Mehler, J., & Nespor, M. (2004). Linguistic rhythm and the development of language. In A. Belletti & L. Rizzi (Eds.), *Structures and beyond: The cartography of syntactic structures* (pp. 213–221). Oxford: Oxford University Press.

Mehler, J., Peña, M., Nespor, M., & Bonatti, L.L. (2006). The 'soul' of language does not use statistics: reflections on vowels and consonants. *Cortex*, **42**, 846–854.

Mervis, G.B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 201–233). Cambridge: Cambridge University Press.

Miller, G.A. (1996). *The science of words* (2nd edn.). New York: W.H. Freeman.

Morgan, J.L., & Demuth, K. (Eds.) (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Murdock, G.P. (1959). Cross language parallels in parental kin terms. *Anthropological Linguistics*, **1**, 1–5.

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: differences between consonants and vowels. *Cognition*, **98**, 13–30.

Nazzi, T., & Bertoncini, J. (2009). Phonetic specificity in early lexical acquisition: new evidence from consonants in coda positions. *Language and Speech*, **52**, 463–480.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, **24**, 756–766.

Nazzi, T., Floccia, C., Moquet, B., & Butler, J. (2009). Bias for consonantal over vocalic information in French- and English-learning 30-month-olds: crosslinguistic evidence in early word learning. *Journal of Experimental Child Psychology*, **102**, 522–537.

Nazzi, T., & New, B. (2007). Beyond stop consonants: consonantal specificity in early lexical acquisition. *Cognitive Development*, **22**, 271–279.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue & Linguaggio*, **2**, 201–227.

Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable-timed languages. In M. van Oostendorp, C.J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology*. 5 vols. Malden, MA and Oxford: Wiley-Blackwell.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realization in VO and OV languages. *Lingue e Linguaggio*, **7**, 1–28.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Berlin: Mouton de Gruyter. (2nd edn., 2008. Dordrecht: Foris.)

New, B., Araujo, V., & Nazzi, T. (2008). Differential processing of consonants and vowels in lexical access through reading. *Psychological Science*, **19**, 1223–1227.

Newport, E.L., & Aslin, R.N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, **48**, 127–162.

Pisoni, D.B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, **13**, 253–260.

Pons, F., & Toro, J.M. (2010). Structural generalizations over consonants and vowels in 11-month-old infants. *Cognition*, **116**, 361–367.

Prunet, J.F., Beland, R., & Adrissi, A. (2000). The mental representation of Semitic words. *Linguistic Inquiry*, **31**, 609–648.

Ramus, F., Hauser, M.D., Miller, C.T., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and cotton-top tamarin monkeys. *Science*, **288**, 349–351.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, **73**, 265–292.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926–1928.

Saffran, J.R., Pollak, S.D., Seibel, R.L., & Shkolnik, A. (2007). Dog is a dog is a dog: infant rule learning is not specific to language. *Cognition*, **105**, 669–680.

Selkirk, E.O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Shi, R., Morgan, J., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language*, **25**, 169–201.

Shukla, M., & Nespor, M. (2010). Rhythmic patterns cue word order. In N. Erteschik-Shir (Ed.), *The sound patterns of syntax* (pp. 174–188). Oxford: Oxford University Press.

Soderstrom, M., Seidl, A., Kemler Nelson, D., & Jusczyk, P. (2003). The prosodic bootstrapping of phrases: evidence from prelinguistic infants. *Journal of Memory and Language*, **49**, 249–267.

Soja, N., Carey, S., & Spelke, E. (1985). Constraints on word learning. Paper presented at the 1985 Biennial Convention of the Society for Research in Child Development, Toronto, Canada.

Stager, C.L., & Werker, J.F. (1997). Infants listen for more phonetic detail in speech perception than in word learning tasks. *Nature*, **388**, 381–382.

Toro, J.M., Nespor, M., Mehler, J., & Bonatti, L.L. (2008a). Finding words and rules in a speech stream: functional differences between vowels and consonants. *Psychological Science*, **19**, 137–144.

Toro, J.M., Shukla, M., Nespor, M., & Endress, A.D. (2008b). The quest for generalizations over consonants: asymmetries between consonants and vowels are not the by-product of acoustic differences. *Perception and Psychophysics*, **70**, 1515–1525.

Werker, J.F., & Tees, R.C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, **7**, 49–63.

Received: 30 August 2010
Accepted: 27 June 2011

Supporting Information

Additional supporting information may be found in the online version of this article.

Figure S1 Percentage of different consonant sequences (in blue) and different vowel sequences (in red) for vocabulary sizes from 10 to 400 words, in Italian.

Figure S2 Percentage of different consonant sequences (in blue) and different vowel sequences (in red) for vocabulary sizes from 10 to 400 words, in French.

Appendix C. Interference effects over memory for music and speech

The experiments 3 and 4 of this thesis showed that at birth, human infants do not represent all auditory stimuli in the same way. Newborns rather seem to represent sounds of the same nature (e.g. all speech sounds) using similar formats, whereas they represent sounds of different nature (e.g. music and speech) using different representational formats. These results showed that the human brain is prewired to treat music and speech differently, perhaps suggesting that this distinction has a biologically determined function for our species.

The findings of Chapter 3 are of particular interest for the ongoing debate about the foundation, neural representations, and origins of music and speech in the human brain (see Patel, 2003, 2010; Peretz, 2006; Peretz & Coltheart, 2003). Before entering in such debate, we decided to perform ourselves a series of adult experiments that could justify the experiments that we then carried out with newborns. In the first experiment we examined participants' recognition of sentences spoken in unknown languages. We evaluated participant's memory after silent retention intervals, after hearing intervening sentences (also spoken in foreign languages), or after intervening music. In the second experiment, another group of participants was tested in similar conditions, except that recognition of musical excerpts, instead of sentences, was tested. The results of those pilot experiments are presented in this appendix.

Experiment C1: Memory for sentences spoken in unknown languages

Participants

Twenty participants (11 males, mean age 23 years old, range 19-25) were tested individually in a sound attenuated room. They were all right-handed and reported no hearing deficits or neurological/psychiatric diseases. All participants were native speakers of Italian and were naïve about the purposes of the experiments. Criteria of participation included no knowledge of the languages used in the experiments and no formal music training at least in the 5 years previous to the experiments. Individuals received remuneration for their participation.

Stimuli

A hundred and twenty recorded sentences constituted the speech database. There were sentences of Basque, Polish, Finnish, Turkish, Arabic, Dutch, Hungarian and Japanese recorded by three different native speakers of each language. Each sentence lasted 3s, and was edited with Praat (Boersma & Weenik, 2008) to have a mean intensity of 70dB. Additionally, a set of hundred and twenty high-quality music loops constituted the music database. The loops were recorded from guitar, harmonics, bongos, flute, electronic piano, pandeiro, violin, and drums; they were 3s long and were edited in Praat (Boersma & Weenik, 2008) to have 70dB of intensity. Twenty different loops were used in the SMS Condition (see next section).

Procedure

Participants were tested individually in a silent and dimly illuminated room. The experiment consisted of an exposure phase, a retention interval and a test phase. During the exposure phase, participants were instructed to listen to sentences passively. Twenty sentences were presented to each participant, one sentence at the time, followed by a 2s pause. After a silent pause of 10s the same twenty sentences were played a second time to the subjects. The sentences were chosen semi-randomly from the database; we only ensured that no more than 3 sentences of each language were presented in a raw. After the exposure phase, a 2-minutes retention interval began. The retention interval was either silent (Condition S_S), or was filled with a new set of 20 interfering sentences (Condition SSS) or with 20 instrumental music excerpts (Condition SMS). All participants performed the three test conditions in three test sessions. The sequential order of the conditions was counterbalanced across participants. In the test phase of each experiment, participants listened to 40 sentences (twenty sentences from the exposure phase and twenty novel sentences presented in an interleaving and random fashion). The number of familiar sentences spoken in each language matched the number of novel sentences spoken in that language. For example, if there were three sentences spoken in Basque in the exposure phase, there were also three new sentences spoken in Basque in the test phase. Participants were instructed to indicate with a button press whether they recognized the sentence as been presented during the exposure phase or as a totally novel sentence. The sentences and the music were presented using high quality headphones. The experiment was run in a MacBook Pro running the Mac OS X operating system using the PsyScope X software package (http://psy.ck.sissa.it/). The total session lasted approximately 45 minutes.

Data Analysis

Participants accuracy was measured using A' (Pallier, 2002). A' was calculated in the following way:

$$A' = \frac{1}{2} + \frac{(hit - fa) * (1 + hit - fa)}{4 * hit * (1 - fa)}$$

where *hit* corresponds to the number of correct responses (number of trials in which the participant correctly recognize the familiar sentence), *fa* stands for the number of false alarms (number of trials in which the participant wrongly answered that a novel sentences was presented during the exposure phase). A' near 1.0 indicates good discriminability, whereas a value near 0.5 means chance performance.

Results

Participants performed above chance in the three conditions (all $ps < 1x10^{-13}$). A one way ANOVA with Condition (S_S/SSS/SMS) as a within subject variable yielded a main effect of Condition, F (3,17)=4.429; $p < 0.03$. Post-hoc pair t-tests showed that this difference was due to the participants' overall higher performance in the S_S (mean A'= 0.87 ± 0.05) as compared to the SSS condition (mean A' = 0.83 ± 0.06); t(19) = 2.766; $p < 0.01$. Participants also showed a lower performance in the SMS (mean A'= 0.84 ± 0.06) as compared to the S_S condition, however, no significant difference was found between these two conditions t(19) = -1.818; $p = 0.08$. The mean A' scores for each experimental condition is depicted in figure C1.

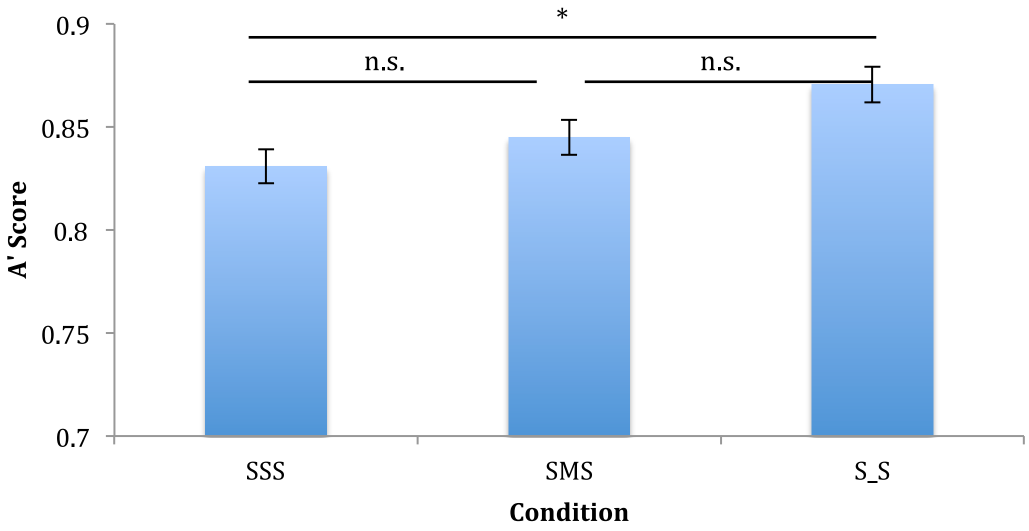

Figure C1. Participants A' scores in the recognition of sentences spoken in unfamiliar languages. S_S: Silent retention interval; SSS: Retention interval filled with interfering sentences; SMS: Retention interval filled with musical excerpts. Error bars represent SEM. There were significant differences between the S_S and the SSS conditions. *, $p<0.01$

Experiment C2: Memory for music excerpts

Participants

Another group of twenty participants (12 females, mean age 24.6 years old, range 19-33y) were tested individually in a sound attenuated room. All participants were native speakers of Italian and were naïve about the purposes of the experiments. They were all right-handed and reported no hearing deficits or neurological/psychiatric diseases. Criteria of participation included no knowledge of the languages used in the experiments and no formal music training at least in the 5 years previous to the experiments. All participants received remuneration after the experiment.

Stimuli

The stimuli were extracted from the same databases used in the previous experiment.

Procedure

The procedure was identical to the one used in the previous experiment, except that instead of sentences, participants listened to musical excerpts during the exposure phase. Twenty different excerpts were presented to each participant, one excerpt at the time, followed by a 2s pause. After a silent pause of 10s the same twenty excerpts were played a second time to the subjects. The excerpts were chosen semi-randomly from the database; we only ensured that no more than 3 excerpts of the same instrument were presented in a raw. After the exposure phase a 2-minutes retention interval began. The retention interval was either silent (Condition M_M), or contained a set of twenty interfering sentences (Condition MSM) or twenty novel excerpts (Condition MMM). All participants performed the three test conditions in three separated tests. The sequential order of the conditions was counterbalanced across participants. In the test phase participants listened to forty excerpts (the same 20 excerpts of the exposure phase and 20 new excerpts presented in an interleave, random fashion). Participants were instructed to indicate with a button press whether they recognized the excerpt as been presented during the exposure phase or as a totally novel excerpt.

Data Analysis

The analysis was identical to the one performed in the previous experiment.

Results

Participants performed above chance in the three conditions (all *ps* < $1x10^{-15}$). A one way ANOVA with Condition (M_M/MMM/MSM) as within subject variable yielded a main effect of Condition, F (3,17)=4.166; $p < 0.05$. Post-hoc pair t-tests showed a significant difference between the M_M (mean A'= 0.88 ± 0.06) and the MMM conditions (mean A' = 0.82 ± 0.10); t(19) = 2.736; $p < 0.05$. Moreover, a significant difference was found between the MSM (mean A'= 0.88 + 0.07) and the MMM conditions; t(19) = 2.387; $p < 0.05$. The mean A' scores of each experimental condition is depicted in figure C2

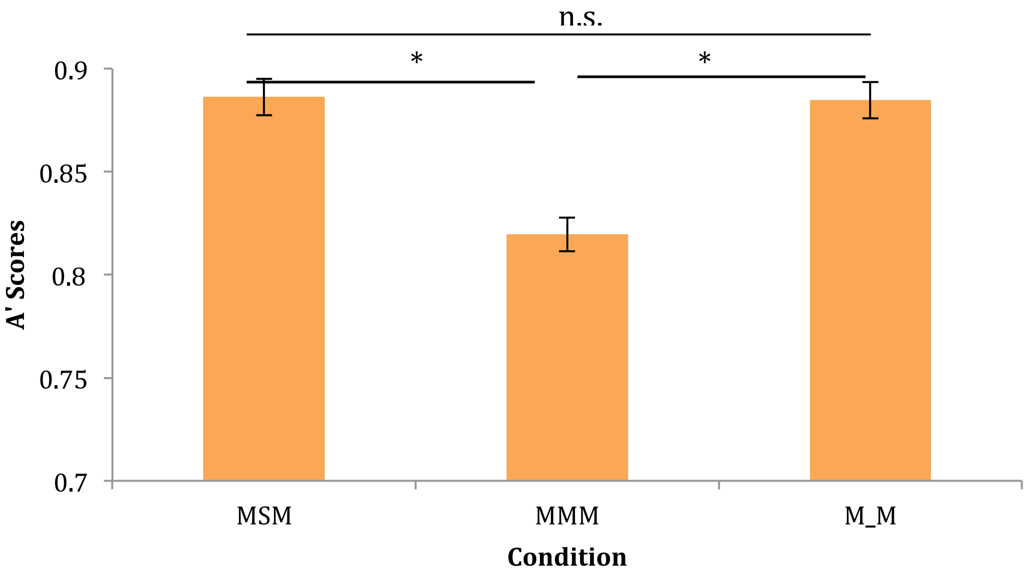

Figure C2 Participants A' scores in the recognition of excerpts. M_M: Silent retention interval; MSM: Retention interval filled with interfering sentences spoken in unfamiliar languages; MMM: Retention interval filled with novel musical excerpts. Error bars represent SEM. There were significant differences between the MMM and the MSM and M_M conditions. *, p<0.05

Discussion

The two pilot experiments carried out with adults suggested that interference effects are stronger when participants listen to interfering sounds of similar nature. In the first experiment, when participants had to recognize sentences, they were less accurate after listening to other sentences than after a silent retention interval. No such effect was observed when participants listened to interfering music; in fact, the participants' performance in the SMS condition was not significantly different from their performance in the baseline condition.

In the second experiment when participants had to recognize excerpts, they were less accurate after listening to other excerpts than in the baseline condition (in which no sound was presented during the retention interval); however their performance was not different from the baseline after listening to interfering sentences. These results suggested that music and speech might use different representational systems, or might be represented separately in the human brain.

Appendix D. Newborns' memory for instrumental music

Speech and instrumental music are sounds that are frequently present in the typical environment of a newborn infant. Previous studies suggest that early in life and until adulthood, the human brain processes speech and music in distinct neural networks (Zatorre, Belin, & Penhune, 2002; Peretz & Coltheart, 2003; Dehaene-Lambertz et al., 2010; Perani et al., 2010). Moreover, our own research (see Chapter 3) shows that since birth, the human brain stores words in a format that do not overlap with the representation of music stimuli. In this appendix we present our studies on newborn's memory for instrumental music. The aim of the studies is to assess similarities and differences in the hemodynamic responses when newborns recognize previously heard instrumental music as compare to memory for word sounds. Additionally, we attempt to explore the cortical brain structures that are involved in the recognition of musical excerpts at birth. In the first experiment of this appendix we tested newborn's memory for a 10s long melody (presented once in each block), whereas in the second experiment we explored the newborn's ability to recognize short excerpts (repeated 4 times within a block).

Experiment D1: Newborn's memory for melodies

Participants

Thirty-eight neonates (20 females, mean age 2.9 days, range 2-4 days) were included in the analysis. Six additional neonates were tested but excluded from the analysis because of head movements that produced large motion artifacts (n=4), because they cried before the end of the experiment (n=1), or because of experimental error (n=1). Neonates were recruited at the newborn nursery at the Azienda Ospedaliera Universitaria Santa Maria della Misericordia in Udine, Italy. Participants were considered eligible if they had gestational ages between 38 and 42 weeks, Apgar scores ≥8 in the firsts minutes, diameter of head ≥33.5cm and uncomplicated history during pregnancy and delivery. Moreover, we did not include infants' whose parents played the piano or heard instrumental music during pregnancy. Bioethics Committee of SISSA/ISAS (International School for Advanced Studies) approved the study; all parents signed an informed consent before the experiments.

Stimuli

Two simplified piano arrangements (Valse Op. 39 No.3 and Valse Op.39 No.15) were prepared in a MIDI format (See figures D1 and D2). The two arrangements had the same tempo and lasted 10 seconds. Moreover they were edited to have the same mean intensity (70dB). The arrangements were prepared at the BRAMS institute in Canada.
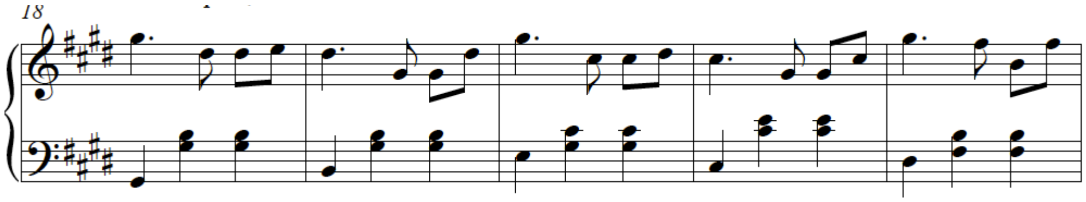
Figure D1. Piano arrangement of the Valse Op.39 No.3 used in the study
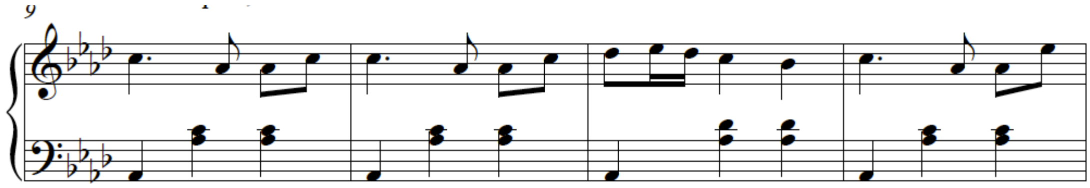
Figure D2. Piano arrangement of the Valse Op.39 No.15 used in the study

Procedure

The experiment consisted of a familiarization phase, a silent interval, and a test phase. The familiarization phase lasted 6 minutes and was organized in 10 blocks. Each block contained a single melody that lasted 10s each. Blocks were separated by time intervals of varying duration (25s or 35s) to avoid synchronization between stimuli occurrences and spontaneous oscillations. A 2-minute silent interval was inserted between the end of the familiarization and the beginning of the test phase. The test lasted for 3 minutes and it consisted of five blocks. In the test, half of the infants heard the same melody that they heard in the familiarization (Same-melody group), whereas the other half was presented with a novel melody (Novel-melody group). The melodies used in the familiarization were counterbalanced across participants. All analyses used pooled data from all newborns, since there was no significant difference between participants in any of the blocks (permutation test, all *ps* > 0.24), evidencing that the acoustic properties of the excerpts did not produce different neural responses that could be captured in the analysis.

As in the experiments presented in Chapters 2 and 3, neonates were tested while lying in their cribs, asleep or in a state of quiet rest. A medical doctor assisted neonates inside the sound-attenuated booth where the experiment was run. The music was presented via two loudspeakers placed at a distance of 1.2m from the infant's head at a 30° angle on both sides, raised to the same height as the crib. The speakers were connected to a Macintosh power PC G5 computer that at the same time operated the NIRS machine and presented the auditory stimuli using PsyScope X software (http://psy.ck.sissa.it/). Both the NIRS machine and the computer were placed outside the experimental booth and were controlled by the experimenter. An infrared video camera was used to monitor the infant's behavior.

Data acquisition

The placement of the probes and data acquisition was identical to the one used in the Chapters 2 and 3.

Data processing and analysis

To assess whether the two conditions differed across the familiarization phase, we computed the maximum difference between the mean activation for each condition. We evaluated significance using non-parametric methods. Specifically, we used permutation tests (Nichols & Holmes, 2002). In these tests, a distribution for the test statistic under the null hypothesis is obtained by re-randomizing the condition assigned to each subject. Significance is then computed as the proportion of reassignments exhibiting a value greater or equal than the one associated to the original groups. Other statistics are built and statistically evaluated in the same way. We used 10,000 random reassignments for each permutation test.

Results

Averaging across all channels and infants per group, we found no block showing significant differences between the groups (Figure D3). Qualitatively we observed an increment of oxy-Hb in the Same-melody groups and a decrement in the Novel-melody groups in the second and third blocks of the test phase. However we found no significant differences in brain activation between the groups in any of these blocks (second block of the test: p=0.20; third block of test p=0.25).

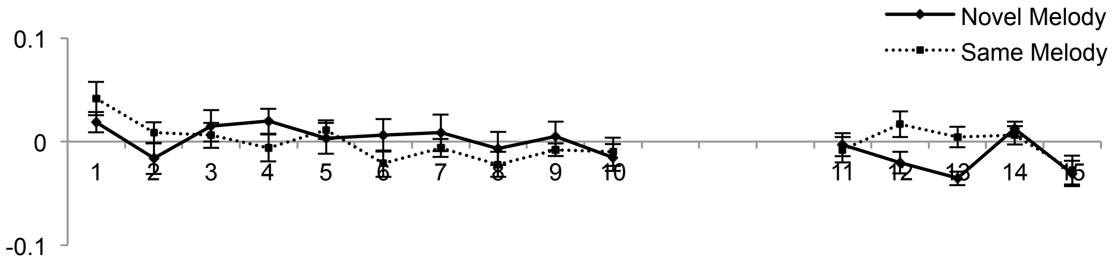

Figure D3. Memory for melodies: time courses of the relative hemodynamic changes. Data are averaged across all the channels and subjects per group. There were no significant differences between the conditions in the test phase. The dashed line indicates the time series for the group that heard the same melody before and after the pause; the continued line represents the group that heard a novel melody in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed.

We also computed permutation tests in six different areas of the brain (frontal, temporal and parietal regions of both hemispheres). We found significant differences between the groups in the third block of the test phase in the left frontal (permutation test, $p<0.01$; see figure D4), left temporal (permutation test, $p<0.01$; see figure D5), and left parietal areas (permutation test, $p<0.02$; see figure D6). No other block showed significant differences in brain activity between neonates in the Novel-melody and in the Same-melody groups, in either the familiarization or test phases.

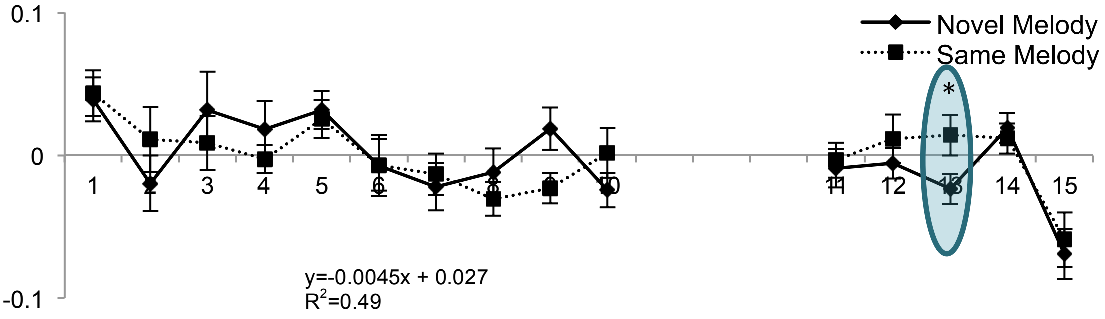

Figure D4. Memory for melodies: Time courses of the relative hemodynamic changes measured in the ***left frontal area.*** There was a significant reduction in the response throughout the experiment. Moreover, there was a significant difference between the groups in the third block of the test phase. The dashed line indicates the time series for the group that heard the same melody before and after the pause. The continued line represents the group that heard a novel melody in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed. *, p <0.01.

Moreover, a linear regression analysis computed over the time course of the experiment showed a significant reduction in the hemodynamic response in the left anterior area ($y=-0.0045x + 0.027$; coefficient 1 $p = 0.03$; coefficient 2 $p = 0.03$; Novel-Melody $R^2=0.2516$; Same-Melody $R^2=0.5265$; all infants $R^2=0.490$. For statistical purposes, we also compared the beginning (first five blocks) and the end (last five blocks) of the familiarization phase using t-test. This comparison further showed that the hemodynamic response changed throughout the experiment in the left frontal areas, t(4)=3.881, *p*<0.01.

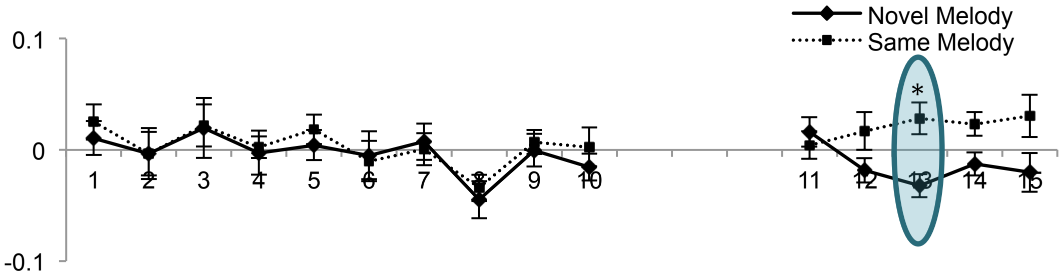

Figure D5 Memory for melodies: Time courses of the relative hemodynamic changes measured in the ***left temporal area.*** There was a significant difference between the groups in the third block of the test phase. The dashed line indicates the time series for the group that heard the same melody before and after the pause. The continued line represents the group that heard a novel melody in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed. *, p <0.01.

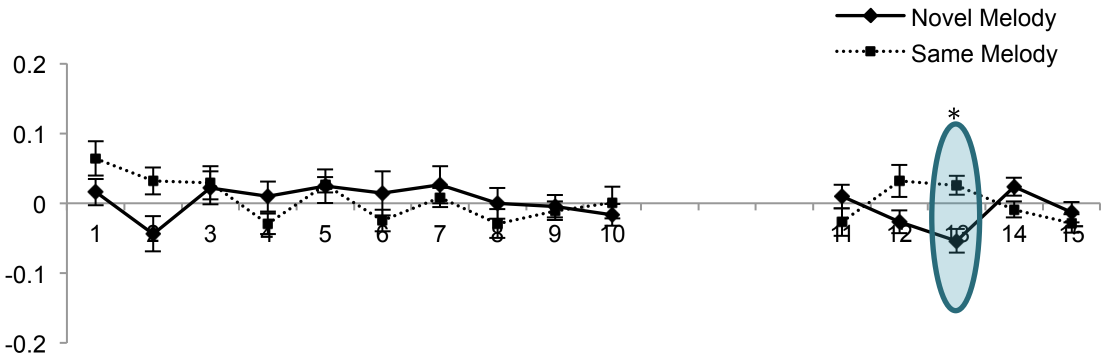

Figure D6 Memory for melodies: Time courses of the relative hemodynamic changes measured in the ***left parietal area.*** There was a significant difference between the groups in the third block of the test phase. The dashed line indicates the time series for the group that heard the same melody before and after the pause. The continued line represents the group that heard a novel melody in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed*, p <0.02.

Discussion

These results show that in similar paradigms the neonate brain reacts differently when presented with melodies and when presented with words. In particular, when newborns heard a familiar excerpt we observed an increment in oxy-Hb that contrasts with the decrement in the response observed when newborns listened to a familiar word (see Chapters 2 and 3). Additionally in this experiment, when newborns were presented with a novel excerpt, there was decrement in the hemodynamic response whereas in the experiments of memory for words we found an increment in oxy-Hb when a novel word was presented in the participants. The pattern of response in the present study also differed from the response found in previous fNIRS studies in its timing. In this study, significant differences between the groups were observed in the third block of the test phase, whereas in the word recognition experiments the significant differences were found in the first block of the test phase. Still, the significant differences between the groups suggest that some traces of the excerpts were kept in the newborns' brains.

There may be two possible explanations for the differences found in the pattern of response. First, the differences might be due to the intrinsic properties of the stimuli (music and speech may recruit different neural circuits at birth). Second, the differences might arise because of the differences in the amount of exposure to each stimulus that infants receive during the familiarization phase. Indeed in the current experiment newborns heard 10 times each melody during the familiarization, whereas in previous experiments newborns heard 60 repetitions of the bisyllabic words. The reduced exposure in the current experiment could have thus delayed the recognition response in the test, or produce and inverted response. The following experiment was carried out to investigate these alternatives.

Experiment D2: Newborn's memory for short motifs

Participants

Thirty-eight neonates (18 males, mean age 3.0 days, range 2-4 days) were included in the analysis. Seven additional neonates were tested but excluded from the analysis because of head movements that produced large motion artifacts (n=3), or because they cried before the end of the experiment (n=4). Neonates were recruited at the newborn nursery at the Azienda Ospedaliera Universitaria Santa Maria della Misericordia in Udine, Italy. Participants were considered eligible if they had gestational ages between 38 and 42 weeks, Apgar scores ≥8 in the firsts minutes, diameter of head ≥33.5cm and uncomplicated history during pregnancy and delivery. Moreover, only infants' whose parents did not play the piano or hear instrumental music during pregnancy were accepted for the test. Bioethics Committee of SISSA/ISAS (International School for Advanced Studies) approved the study; all parents signed an informed consent before the experiments.

Stimuli

Two simplified motifs were prepared in a MIDI format. Each motif was composed of 4 piano notes. The motifs had the same tempo and lasted 1.8 seconds each. The length of the notes was the same in the two motifs whereas the tones were different. The motifs were edited to have the same mean intensity (70dB). The arrangements were prepared at the BRAMS institute in Canada.

Procedure

The experiment consisted of a familiarization phase, a silent interval, and a test phase. The familiarization phase lasted 6 minutes and was organized in 10 blocks. Each block contained four identical motifs separated by random pauses of 0.5s, 1s or 1.5s yielding blocks of about 10s each. Blocks were separated by time intervals of varying duration (25s or 35s) to avoid synchronization between stimuli occurrences and spontaneous oscillations. A 2-minute silent interval was inserted between the end of the familiarization and the beginning of the test phase. The test lasted for 3 minutes and it consisted of five blocks. In the test, half of the infants heard the same motif that they heard in the familiarization (Same-motif group), whereas the other half was presented with a novel motif (Novel-motif group). The motifs used in the familiarization were counterbalanced across participants. The testing procedure was identical to the procedure used in the previous experiment.

Data acquisition

The placement of the probes and data acquisition was identical to the ones used in the Chapters 2 and 3.

Data processing and analysis

The analyses performed were identical to those conducted for previous experiments.

Results

Averaging across all channels and infants per group we found significant differences between the groups in the third block of the test phase (p=0.02). No other block in the familiarization or in the test phase showed significant differences between the groups (all *p*s > 0.05). Moreover, we observed an increment of oxy-Hb in the Novel-motif group and a decrement in the Same-motif group in the third block of the test phase. Figure D7 shows the pattern of hemodynamic activity in the experiment.

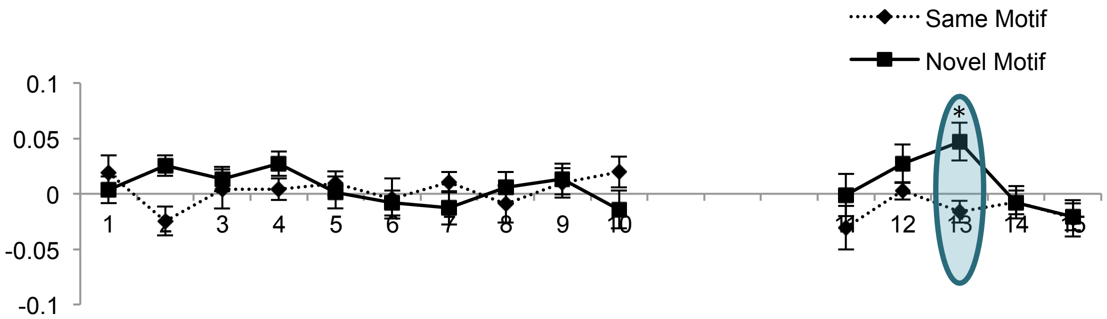

Figure D7. Memory for motifs: Time course of the relative hemodynamic changes ***averaged across all the channels and subjects per group.*** There was a significant difference between the groups in the third block of the test phase. The dashed line indicates the time series for the group that heard the same word before and after the pause; the continued line represents the group that heard a novel word in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed. *, $p<0.05$

The same pattern of activation (increment of oxyHb in Novel-motif group and decrement in the Same-motif group in the third block of the test phase) was observed in a distributed network in the brain. However the right-frontal region (see figure D8) was the only area that independently showed significant differences (*p*<0.05)

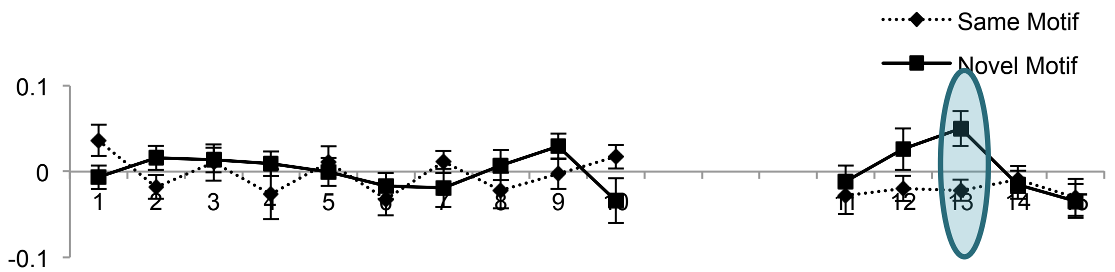

Figure D8. Memory for motifs: Time course of the relative hemodynamic changes in the ***right frontal area.*** There was a significant difference between the groups in the third block of the test phase. Data is averaged across all the channels and subjects per group. The dashed line indicates the time series for the group that heard the same word before and after the pause; the continued line represents the group that heard a novel word in the test. Error bars indicate standard errors. The x-axis shows number of blocks; in the y-axis the changes in concentration of Oxy-hemoglobin in mmol*mm is displayed.

Discussion

Experiment D1 showed that music and speech stimuli elicit different recognition responses in the newborn's brain. The results of the present experiment suggest that these different patterns were in part due to the nature of the stimuli used (music as opposed to speech) and in part due to the differences in the number of repetitions of the sounds that newborns heard during the familiarization.

This experiment verified that, the recognition response of musical stimuli in the newborns' brains is delayed with respect to the recognition response of single words. In the experiments in which memory for words was tested, we found the recognition response in the first block of the test phase. This pattern of activation was replicated throughout the experiments of the thesis. However, in the two experiments of this appendix, in which memory for musical stimuli was tested, the recognition response was found in the third block of the test phase. This suggests that the delay in the response found in experiments D1 and D2 with respect to the experiments in Chapters 2 and 3, is due to the nature of the stimuli.

On the other hand, in this experiment we presented a shorter instrumental music excerpt than the melody used in experiment D1. By having a shorter melody we were able to repeat the stimuli many more times than in experiment D1 while keeping the nature of the sound. The increment in the number of repetitions made the hemodynamic response more similar to the experiments in which we tested memory for words. In fact, the novel sounds elicited a greater oxy-Hb than the familiar sounds just as it was observed in the experiments of Chapters 2 and 3. The greater number of repetitions possibly allowed the newborns' brains to create a different, perhaps more robust representation of the sound, inverting in this way the pattern of the response. Similar inversions of the responses are often found in developmental studies using other types of measures. For instance, in looking time experiments infants show preference for novel elements in certain situations (for instance after long habituation sessions) and familiarity preference in other circumstances (e.g. Fagan, 1984). These changes in the pattern of responses might be determined by the strength of the representation that infants formed during the familiarization (Jusczyk, 1995). In the same way, we might hypothesize that, the difference in the number of repetitions in the familiarization of experiments D1 and D2 could have given rise to memories of different strengths and thus this differences may have been reflected in the inverted patterns of the neural responses. Future studies, for instance using longer speech stimuli (e.g. short phrases), should be able to verify this hypothesis.

Appendix E. Evaluation of the switch technique to study word learning in 7 month-olds

This appendix presents a pilot study that preceded the selection of the technique used in Chapter 4. Different from the experiments presented in that Chapter (in which we used the interleave paradigm), here we tested infants using the switch paradigm (see Kovacs & Mehler, 2009b). In such paradigm monolingual infants easily learned to associate a linguistic structure to one side of a screen, but required more time to switch to the other side of the screen that was associated to a different linguistic structure, namely they had difficulties overcoming the response that they had learned at first.

In this study we presented a single word (instead of a structure) and evaluate the infants' ability to associate this word to puppets appearing in one side of the screen (left or right). After sixteen trials, infants heard either a totally new word, or a word which middle syllable differed from the previous word; in both cases the change of the word is accompanied by a switch in the location of the puppets (i.e. if the puppets had initially appeared in the right side of the screen, they would appear in the left side after the word changes). We hypothesized that the more similar the representation of the two words (pre-switch and post-switch word) is, the more difficult the task should be for the infant.

Participants

Participants were twenty-eight infants. Half of the participants was assigned to the Middle-change group (9 females); aged from 6 months 25 days to 7 months 15 days (mean age 7 months 7 days). The other half was assigned to the Different-word group (9 females); aged from 6 months 15 days to 7 months 30 days (mean age 7 months 10 days). All participants were monolinguals, with no hearing or visual problems reported. Two additional infants were excluded from the analysis because of crying or fussiness. Parents of the infants signed the informed consent after they had understood the procedure and after all their questions had been answered. The Ethics Committee of the Scuola Internazionale Superiore di Studi Avanzati approved the study.

Stimuli

Infants were presented with trisyllabic CVCVCV words. In the pre-switch phase half of the infants heard the nonce word *pizole*, whereas the other half was presented with the nonce word *mavefi.* In the post-switch phase infants who were assigned to the Middle-change group heard either the word *pivale* or the word *marufi;* infants assigned to the Different-word group heard one of the following words: *gunimo, monide, morisa, desake, sanide, degusa, gusade, serigu, nikesa, sudeni, kesamo, rikegu.* The stimuli were synthesized with the female voice of the MBROLA Italian database IT4 (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996). The duration of each syllable was 200ms and a monotonous pitch of 240Hz was used. There were no pauses between syllables.

The visual stimuli consisted of one central attractor and colorful puppets presented inside two white squares (one presented at the time). The two white squares had a side-length of 8cm, with a distance between them of 13.5 cm. The colorful puppets were used as visual reinforcement. Three puppets were randomly paired with the words in the pre and post-switch phases.

Procedure

The study consisted of 16 pre-switch and 16 post-switch trials. Trials in the pre- and post-switch phases started with a display of 2 white squares, one on either side of the screen, and a central visual attractor. When the infant fixated the central attractor, one word was played. When the word ended, only the 2 white squares were visible for 1 s. Then a looming puppet accompanied by a tinkling sound appeared consistently on one side of the screen in the pre-switch phase and on the other side of the screen in the post-switch phase. The initial side was counterbalanced across participants.

Data acquisition

Infants' gaze was recorded with an eye-tracker (TOBII 1750). The eye tracker was integrated into a 17-inch TFT screen, where the stimuli were presented via an IMAC 10,1 running PsyScope X software (http://psy.ck.sissa.it/). A loudspeaker was placed behind the screen for the presentation of the acoustic stimuli. Infants were seated on their parent's lap at about 50 cm distance from the monitor. A hidden videocamera was used to observe the infant's behavior.

Data analysis

We measured the infant's first fixation after hearing the test word and before the beginning of the next trial. We divided the screen into three equal parts: left, middle and right. Trials in which the infant performed an anticipatory look to the side (left or right) where the puppet would appear were coded as correct. If the infant did not look to the correct side during the anticipatory period the trial was coded as incorrect.

Results

Figure E1 shows the results pooling data of all infants. In figures E2 and E3 data of the two groups (Middle-change and Different-word) are presented separately. We performed an ANOVA using the Groups as between-subjects factor, phases (pre and post-switch) and trials (early and late trials) as within-subjects factors. We found a main effect of phase $F(1,14) = 58.435$, $p < 0.0001$, due to the fact that overall, infants were significantly more correct in the pre-switch than in the post-switch phase (Middle-change group $t(15)=6.455$, $p=1\text{x}10^{-5}$; Different-word group $t(15)=6.220$, $p=1\text{x}10^{-5}$; all infants $t(15)=7.395$, $p=2\text{x}10^{-6}$). This suggested that infants had greater difficulties to learn in the post-switch phase than in the pre-switch phase.

We also found a main effect of trial $F(1,14) = 16.313$, $p < 0.001$, due to the fact that infants showed more correct responses in late trials than in early trials. In fact a linear regression analysis computed over the proportion of infants displaying correct anticipatory looks, showed a tendency for more correct anticipations in later familiarization trials both in the pre-switch ($R^2=0.3944$, $t(15)=-2.493$, $p=0.02$) and the post-switch phases ($R^2=0.4450$, $t(15)=-3.967$, $p=0.001$). Finally we found no main effect of group $F(1,14) = 3.089$, $p > 0.05$, but there was a significant interaction trial*phase*group $F(1,14) = 6.140$, $p < 0.05$.

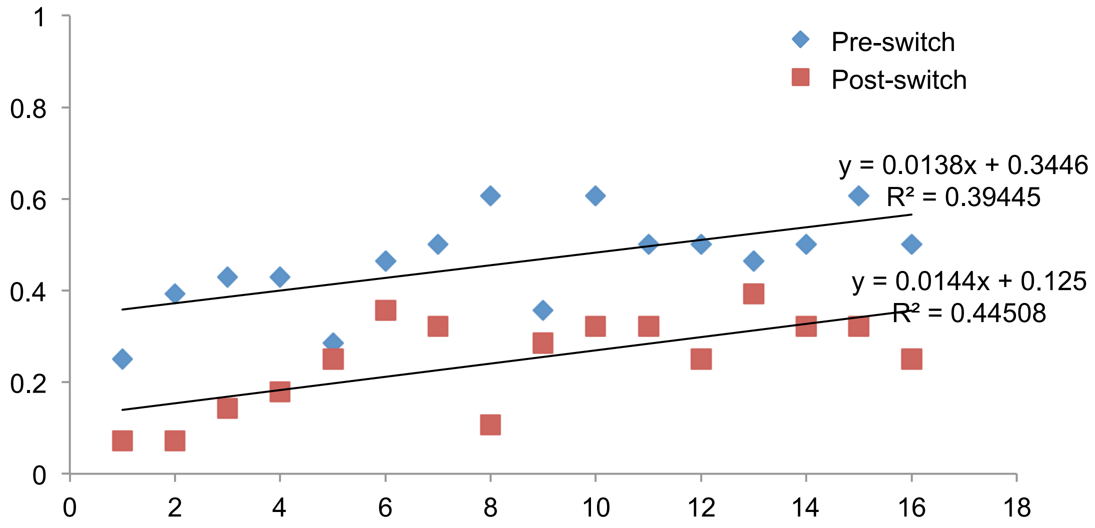

Figure E1. Data pooled across participants: Proportion of infants with correct anticipatory looks. Blue diamonds represent population averages for the pre-switch phase, and red dots represent averages for post-switch phase. Regression lines are shown for both groups. Participants increased anticipatory looks to the correct side in both phases, however infants showed significantly more correct anticipatory looks in the pre-switch than in the post-switch phase.

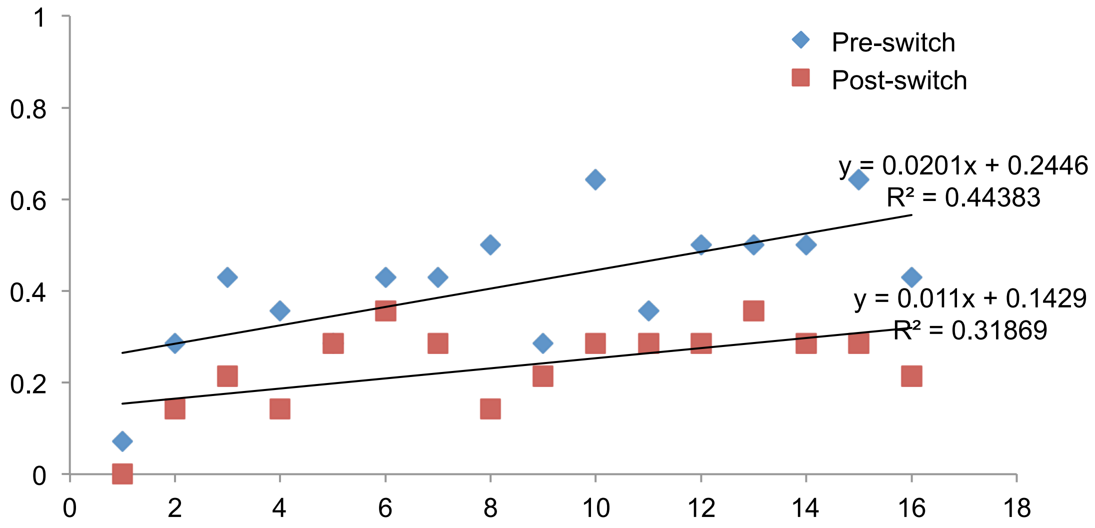

Figure E2. Proportion of infants in the Middle-change group with correct anticipatory looks. Blue diamonds represent population averages for the pre-switch phase, and red dots represent averages for post-switch phase.

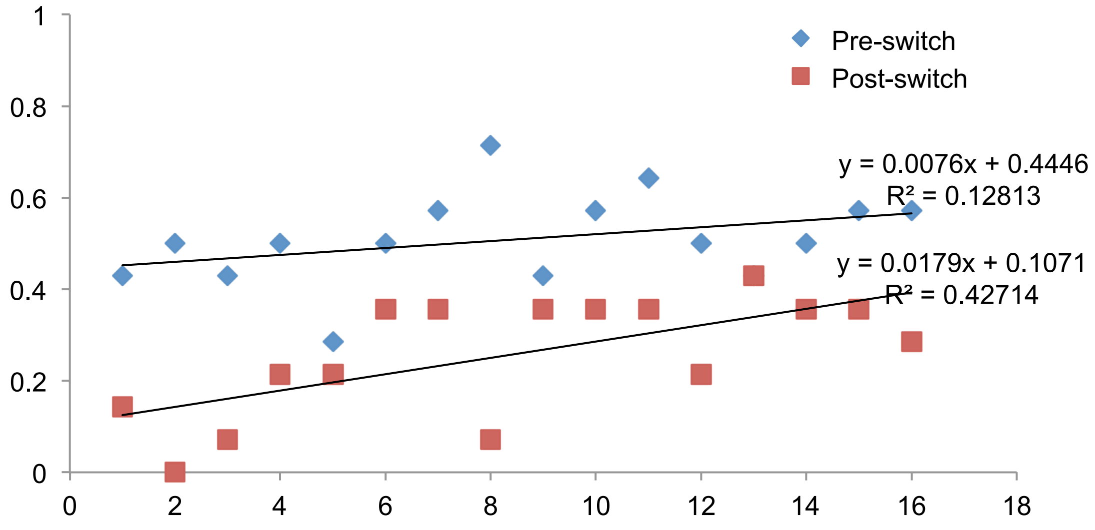

Figure E3. Proportion of infants in the Different-word group with correct anticipatory looks. Blue diamonds represent population averages for the pre-switch phase, and red dots represent averages for post-switch phase. Regression lines are shown for both groups.

Discussion

One the one hand, these results confirmed the original study of Kovacs and Mehler (2009a) showing that monolingual infants have greater difficulties to learn a second response (association of a sound to one side of the screen) than the first association. This difference in performance between the pre and post-switch phases is likely to arise because infants have difficulties inhibiting the first response. On the other hand the data did not confirm our initial hypothesis: independently of how similar (or different) the pre and post words were, monolinguals seemed to have the same difficulties in the task. We concluded that this technique was not sensible to test word learning in 7 month-olds.

Appendix F. Auditory memory span in 10 month-olds

How many elements can young infants retain in working memory? Different experiments suggest that, at least in the visual domain, the capacity limit is 3 items. The evidence comes from at least two different paradigms used in developmental literature. In one investigation, 10- and 12-month-old infants watched an experimenter sequentially placing a number of crackers into one bucket (e.g. 2 crackers) and a different number of crackers into another bucket (e.g. 3 crackers). Infants were then allowed to approach the containers. The study showed that infants chose first the bucket containing the greater quantity when 1,2, or 3 crackers were hidden in either buckets. However, when the experimenter hid more than 3 crackers in either bucket, infants chose randomly (Feigenson & Carey, 2005; Feigenson, Carey & Hauser, 2002). This pattern suggested that infants remember up to 3 objects per location, but no more. In addition, the results suggested that infants presented with more than 3 objects exhibited a *catastrophic memory failure*, such that they appeared to remember none of the objects in the array (instead of remembering a subset, for instance the first, or first two objects).

Another paradigm used to test working memory span in young infants is the manual search task. Feigenson and Carey (2003, 2005) presented 12- to 14-month-old infants with varying numbers of identical objects simultaneously visible. Infants saw the experimenter hide these objects in an opaque box, and then were allowed to retrieve all or just a subset of them. By measuring the amount of time infants spent searching for the hidden objects when the full set of objects was retrieved as compared to the amount of time spent when only some of the objects had been retrieved, the authors were able to measure how many objects infants retained in working memory. This method revealed that young infants searched appropriately when 1,2, or 3 objects were hidden (e.g. after seeing 3 objects hidden and retrieving just 2 of them, infants continued searching the box for the missing object). However, when infants saw 4 objects hidden and only retrieved 1 or 2 of them, they again showed the catastrophic memory failure, that is they failed to search for any remaining objects, independently of the number objects that were retrieved (1, 2 or 3).

This intriguing phenomenon continues to be subject of investigation in the developmental field, however to our knowledge, the studies on early working memory have only focused on the visual domain. This leaves unresolved the question of whether the item-based capacity limitation is a specific feature of the visual system or whether it is a general constraint of memory during development. In the series of studies presented in this appendix, we explore this question by assessing 10-month-olds' auditory memory span.

One of the most prominent features that distinguishes the visual and the auditory systems is the order in which the elements can be processed. Whereas visual items can be presented sequentially and simultaneously, sounds can only be processed in sequential order. For instance, three identical objects can be presented simultaneously and we can still distinguish that they are three of them; however if we hear three identical sounds at the same time we perceive them as being only one.

The studies described above showed that infants' working memory span in the visual domain is three, both when the objects are presented simultaneously and also when they are presented sequentially. However, due to the fact that infants need to process several sounds (syllables, words) in order to acquire different properties of language, learn words, or begin to understand phrases, it is possible that their memory span might be greater than three in the auditory domain.

In these studies we assess 10-month-olds' memory span using the Kovacs & Mehler (2009). In such paradigm young infants were presented with two regularities, each of them associated to a different side (right or left) of the screen. In our modification of the paradigm we teach infants to look to one side of the screen when a given number of syllables is presented (e.g. 2 syllables), and to look to the other side when another set of syllables (e.g. 3 syllables) is presented. Because the identity of syllables varied from one trial to another, and because both regularities were implemented in the same sets of syllables, we hypothesized that infants' success should depend on their ability to form a representation of the number of sequentially presented syllables. Without the ability to form and maintain two distinct representations of the number of syllables, infants should fail to differentiate the two regularities. However, even if infants were able to distinguish between the two types of sequences, we expect them to show correct responses to only one of them. This is based on the results of previous studies using the same paradigm (Kovacs & Mehler, 2009; Hochmann et al., 2011), which have showed that monolingual infants are unable to associate the two regularities to the corresponding side of the screen, even though they are able to learn the two of them.

Therefore, in order to define the capacity limit, we will test different sets of sequences. We begin testing infants' ability to differentiate 2 syllables versus 3 syllables, because this distinction has been previously verified even in newborn infants (Bijeljac-Babic, Bertoncini & Mehler, 1993).

Experiment F1: Infants' ability to differentiate between sets of 2 and 3 syllables

Participants

Participants were eighteen infants (9 males); aged from 9 months 21 days to 10 months 18 days (mean age 10 months 5 days). All participants were full term, with no birth complications, and no hearing or visual problems reported. Seven infants were excluded from the analysis because of crying or fussiness (n=2), or because of side bias (n=1), or because of experimental error (n=4). Parents of the infants signed the informed consent after they had understood the procedure and after all their questions had been answered. The Ethics Committee of the Scuola Internazionale Superiore di Studi Avanzati approved the study.

Stimuli

Linguistic stimuli were consonant-vowel (CV) syllables repeated in each trial either two or three times. The syllables were synthesized with the female voice of the MBROLA Italian database IT4 (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996). The duration of each syllable was 200ms and a monotonous pitch of 240Hz was used. There were pauses of 250ms between syllables. Twenty different syllables [sa, fi, be, to, ku, ka, le, ni, zo, tu, ba, ve, ri, mo, fu, la, se, pi, gu, do] were used. Each of the syllables was randomly chosen in each trial. Every infant heard a syllable in a maximum of two trials in the whole experiment (one per condition).

The visual stimuli were three pictures of colorful toys. Each appeared inside one of two white squares either on the left or on the right side of the screen. The toys loomed from 4 cm to 7 cm inside the squares for 2s. The squares had a side-length of 8 cm, positioned at a distance of 13.5 cm. All three toys could appear after any familiarization item in the window predicted by that item (either left or right side of the screen).

Procedure

An adaptation of the Kovács & Mehler (2009) procedure was used in this study. The experiment consisted of a familiarization phase of 20 trials and a test phase of 8 trials. Trials in the familiarization phase started with a display of a central visual attractor and two white squares, one on the left and one on the right side of the screen. When the infant fixated the central attractor, a syllable was repeated either two or three times. After the offset of the sequence the central attractor disappeared, leaving only the two white squares visible for 1s. Then a looming puppet appeared on one of the two white squares for two seconds (as predicted by the number of syllables presented). The puppet was accompanied by a tinkling bell of 300ms presented with a delay of 800ms with respect to the onset of the visual stimulus. The presentation of a syllable twice predicted the puppet's appearance in one of the squares, whereas the presentation of a syllable three times predicted the puppet's appearance in the other square. The pairing of sequences with puppets' locations was counterbalanced across participants. Two-syllable and three-syllable sequences were presented in an interleaved pseudo-random order. We ensured that each of the sequences were presented at the most three times in a row.

The test trials were similar to the familiarization trials, except that after each sequence no puppet was displayed in either side of the screen. Each type of sequence was presented four times in the test in a pseudo-random order. The first four trials consisted in two trials of each sequence. Two seconds after the word offset, the next trial started.

Data acquisition

Infants' gaze was recorded with an eye-tracker (TOBII 1750). The eye tracker was integrated into a 17-inch TFT screen, where the stimuli were presented via an IMAC 10,1 running PsyScope X software (http://psy.ck.sissa.it/). A loudspeaker was placed behind the screen for the presentation of the acoustic stimuli. Infants were seated on their parent's lap at about 50 cm distance from the monitor. A hidden videocamera was used to observe the infant's behavior.

Data analysis

We measured the infant's first fixation[9] after hearing the test sequence and before the beginning of the next trial. We divided the screen into three equal parts: left, middle and right. Infants were coded as looking either to the correct or incorrect side of the screen according to the sequence that was presented i.e. two or three syllables. Independently for each sequence, we computed difference scores: (#correct looks - #incorrect looks) ⁄ (#correct looks + #incorrect looks). We then computed a t-test to compare the difference scores to chance level of 0. Significantly positive difference scores would indicate that infants learned the corresponding association with the number of syllables of the sequences and that they differentiate between the two types of sequences.

---

[9] Looking times of 80 ms or more were considered a fixation.

Results

The results over the first fixation data are presented in Figure F1 In the sequences in which the syllable was repeated two times infants' mean difference score was 0.36, which was significantly greater than 0, t(17) 2.60; p < 0.01; d' = 0.589. Twelve infants displayed a positive difference score, four infants a negative difference score, and two a null difference score (binomial test, 2-tails. p = 0.07). For the sequences in which the syllable was repeated three times infants' mean difference score was -0.09, which did not differ from chance, t(17) =-0.536; p = 0.60; d' = -0.12. Seven infants displayed a positive difference score, eight infants a negative difference score, and three a null difference score. This distribution did not differ from chance, p = 0.99.

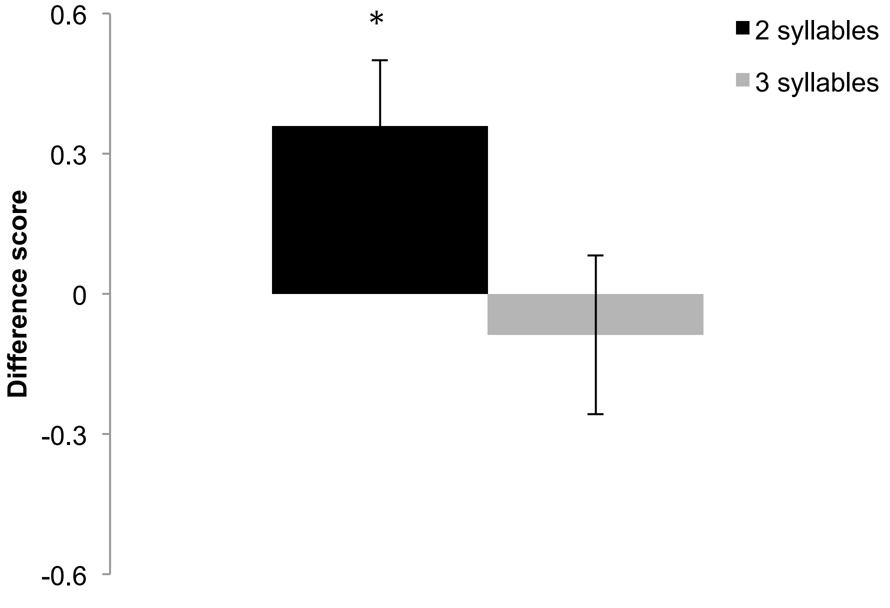

Figure F1. Normalized difference scores considering first fixations. The y-axis shows the mean difference scores. Positive difference scores indicate that infants searched for the puppet in the correct side of the screen whereas negative difference scores indicate that infants searched for the puppet on the incorrect side. Bars depict standard error of the mean (SEM). * p<0.01.

Discussion

These results showed that 10-month-olds were able to associate a two-syllable sequence to the corresponding side of the screen. Moreover, the results suggest that participants were able to differentiate between 2-syllable sequences from the other type of sequences. Presumably, infants were able to differentiate between the two sequences based on the representation of the number of syllables, however, the two types of sequences also differed in length. The 2-syllable sequences lasted 650ms whereas the 3-syllable sequences were 1100ms long. This difference in the duration of the sequences could have determined the infants' performance. The next experiment was carried out to rule out this possibility.

Experiment F2: Infants' ability to differentiate between sets of 2 and 3 syllables of the same length, a control study.

Participants

Participants were twenty infants (10 females); aged from 9 months 17 days to 10 months 19 days (mean age 10 months 2 days). All participants were full term, with no birth complications, and no hearing or visual problems reported. Additional 3 infants were excluded from the analysis because of crying or fussiness. Parents of the infants signed the informed consent after they had understood the procedure and after all their questions had been answered. The Ethics Committee of the Scuola Internazionale Superiore di Studi Avanzati approved the study.

Stimuli

The stimuli were the same used in the previous experiment, except that the sequences were edited in Praat (Boersma & Weenik, 2008) to have the same duration. The algorithm shortened the 3-syllable sequences and lengthened the 2-syllable sequences so that all the sequences lasted 875ms. The visual stimuli were identical to those used in the previous experiment.

Procedure

The procedure was identical to the one used in the previous experiment.

Data analysis

The analyses were identical to those performed in the previous experiment.

Results

The results over the first fixation data are presented in Figure F2. In the sequences in which the syllable was repeated two times infants' mean difference score was 0.36, which was significantly greater than 0, t(19) 2.329; $p < 0.05$; d' = 0.52. Thirteen infants displayed a positive difference score, five infants a negative difference score, and two a null difference score (binomial test, 2-tails $p = 0.09$). For the sequences in which the syllable was repeated three times infants' mean difference score was -0.15, which did not differ from chance, t(19) =-1.195; p = 0.247; d' = -0.27. Four infants displayed a positive difference score, nine infants a negative difference score, and seven a null difference score. This distribution did not differ from chance, p = 0.99.

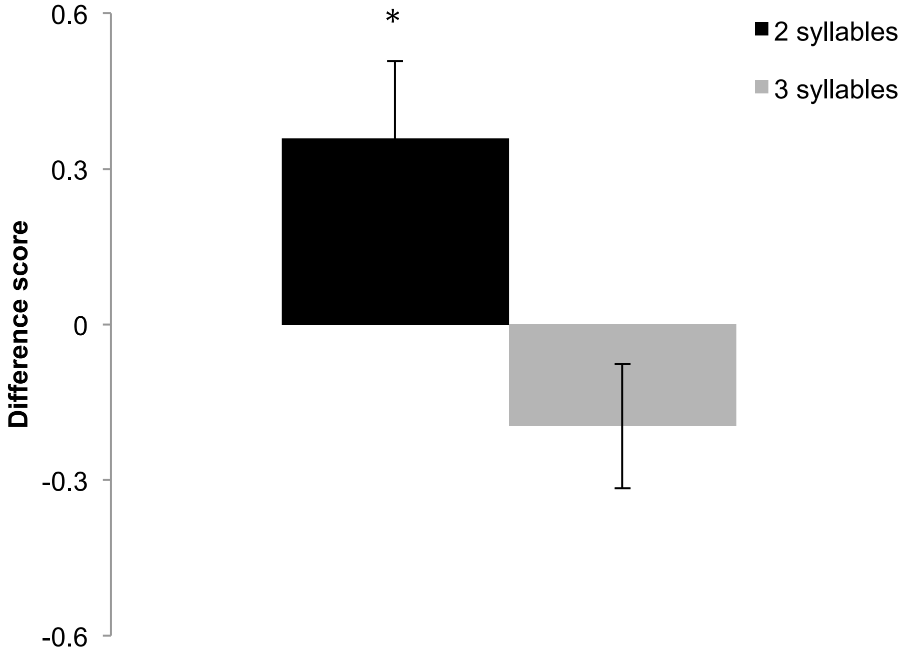

Figure F2. Normalized difference scores considering first fixations. The y-axis shows the mean difference scores. Positive difference scores indicate that infants searched for the puppet in the correct side of the screen whereas negative difference scores indicate that infants searched for the puppet on the incorrect side. Bars depict standard error of the mean (SEM). * p<0.05.

Discussion

This control experiment ruled out the possibility that duration determined infants' performance in the previous experiment. The data rather confirmed the initial interpretation, namely that infants were able to differentiate between two-syllable and three-syllable sequences based on the number of items.

However, we observed a correct association to the corresponding side of the screen only in the 2-syllable sequence. As stated in the introduction, this could reflect the limitations that young infants faced when tested with a paradigm in which they need to associate simultaneously two different regularities to two sides of the screen. Using this paradigm Kovacs & Mehler (2009) showed that monolingual infants displayed correct responses only to one of the two regularities (usually the most prominent or the easiest one), despite of the fact that infants were able to learn both regularities separately. Thus, in order to verify whether infants are able to form a 3-syllable representation, a test in which the 3-syllable sequence becomes the easiest regularity is needed. The next experiment was designed to explore this possibility.

Experiment F3: Infants' ability to differentiate between sets of 3 and 4 syllables.

Participants

Participants were sixteen infants (8 males); aged from 9 months 20 days to 10 months 25 days (mean age 10 months 2 days). All participants were full term, with no birth complications, and no hearing or visual problems reported. Additional 3 infants were excluded from the analysis because of crying or fussiness (n=1), or because of experimental error (n=2). Parents of the infants signed the informed consent after they had understood the procedure and after all their questions had been answered. The Ethics Committee of the Scuola Internazionale Superiore di Studi Avanzati approved the study.

Stimuli

The stimuli were identical to those used in experiment F1.

Procedure

The procedure was identical to that used in previous experiments, except that the syllables were repeated either 3 or 4 times (instead two or three times, as it was done in the previous studies).

Data analysis

The analyses were identical to those performed in the previous experiment.

Results

The results over the first fixation data are presented in Figure F3. In the sequences in which the syllable was repeated three times, infants' mean difference score was 0.14, which was not significantly greater than 0, t(15) 0.774; p = 0.45; d' = 0.19. Nine infants displayed a positive difference score, six infants a negative difference score, and one a null difference score (binomial test 2-tails, p = 0.60). For the sequences in which the syllable was repeated four times infants' mean difference score was -0.19, which did not differ from chance, t(15) =-1.01; p = 0.33; d' = 0.25. Four infants displayed a positive difference score, eight infants a negative difference score, and four a null difference score. This distribution did not differ from chance, p = 0.99.

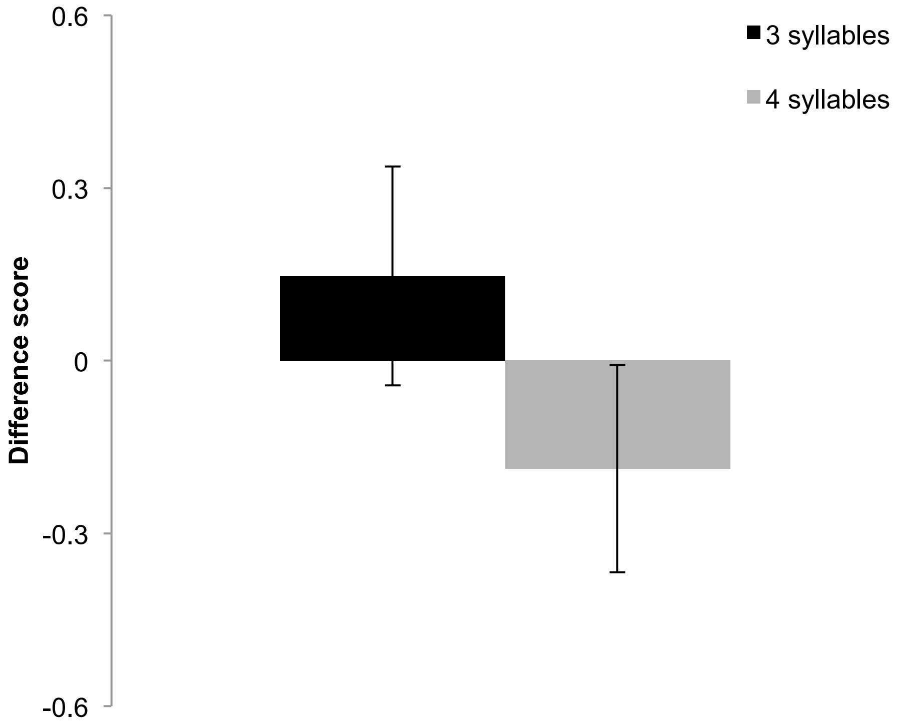

Figure F3. Normalized difference scores considering first fixations. The y-axis shows the mean difference scores. Positive difference scores indicate that infants searched for the puppet in the correct side of the screen whereas negative difference scores indicate that infants searched for the puppet on the incorrect side. Bars depict standard error of the mean (SEM).

Discussion

In Experiments F1 and F2 we observed that 10-month-olds were able to distinguish between two-syllable and three-syllable sequences, however they correctly associated only the 2-syllable sequences to the corresponding side of the screen. The aim of the present experiment was to investigate whether participants can also associate a 3-syllabic sequence to a specific location of the screen. To do that, we thought of testing participants in a paradigm in which they had to differentiate between 3-syllable and 4-syllable sequences (assuming that in this way the 3-syllable sequences would have become the easiest of the two regularities, and therefore that infants would have been able to associate it to the correct side of the screen). However, we did not consider the fact that 4 items could have produced a catastrophic failure to auditory memory, as in visual memory. In fact, the results of the present experiment showed that infants were not able to associate any of the sequences to the corresponding side of the screen. This result might indicate that the span of working memory in the auditory domain is similar to the capacity of working memory in the visual domain. However, before adopting this interpretation we suggest that at least one additional test should be carried out in the future. In such test the infant's ability to associate the 3-syllable sequence to the corresponding side of the screen must be attested. Unfortunately, due to time constraints we were not able to carry out such experiment before the accomplishment of this thesis; however, we suggest that one way to show the infants' ability to associate the 3 syllable sequences to one side of the screen, is by presenting these sequences while sequences of random length (1 or 2 items) are associated to the other side of the screen. We hypothesize that if infants' auditory memory span is 3, they should display the correct responses towards the 3-syllabic sequences in such paradigm, because the three-syllable sequence association would be easier than the random sequence association.

No es nada romántico…decir que tu doctorado fue como un viaje migratorio de tortugas, cuando tú también atravesaste el océano para obtenerlo, pero el mío fue más que eso, así que aquí lo dejo…

…the Costarican turtles travel thousands of kilometer looking for food, but they remember faithfully the very beach where they were born…